



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 99338

TO: Yong D Pak
Location: CM1/10A16/10D01
Art Unit: 1652
Friday, July 25, 2003

Case Serial Number: 10/060848

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Pak,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

1-4 r9 SEQ ID 100:3

2.02.01 60/266,037

- 1) 70%
- 2) 80%
- 3) 90%

a) activity

THIS PAGE BLANK (USPTO)

From: Pak, Yong
Sent: Tuesday, July 22, 2003 7:53 AM
To: STIC-Biotech/ChemLib
Subject: 10/060,848

RECEIVED

JUL 22 2003

(STIC)

dear stic,

please search the following in commercial and interference database for 10/060,848:

1. SEQ ID NO:3
2. oligomer search of SEQ ID NO:3.

thank you.

yong pak
Art Unit 1652

Tel: 703-308-9363
Fax: 703-746-3173
Office: 10A16
Mail: 10D01

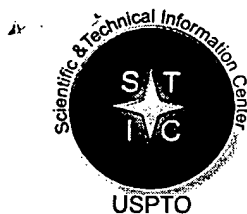
Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 7/24/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: ✓
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ✓
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 – Circ. Desk



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:09:19 ; Search time 99.5 seconds

(without alignments)
954.013 Million cell updates/sec

Title: US-10-060-848-3
Perfect score: 1942
Sequence: 1 MNYSKDAPEVSPKDAREF.....ALBELATKQVQMSYKRLK 366

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriophage:*
17: sp_archaeophages:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	51.9	372	5	062128
2	924	47.6	400	5	09XTB4
3	785	40.4	361	5	081PT9
4	785	40.4	381	5	081PU0
5	785	40.4	446	5	09VP68
6	773	39.8	446	5	0961U7
7	504	26.0	385	5	0917T1
8	480.5	24.7	461	5	08T901
9	472.5	24.3	461	5	09VDP6
10	432	22.2	341	17	08TYB0
11	408	21.0	356	16	08EMJ6
12	407	21.0	369	16	0971J4
13	377.5	19.4	361	2	050353
14	371	19.1	334	16	0930Q4
15	368	18.9	347	16	08NSK9
16	365.5	18.8	366	16	0981V8

17	363.5	18.7	349	16	0828Q4	0828Q4 salmonella
18	361.5	18.6	349	16	08FK55	08FK55 escherichia
19	356.5	18.4	349	16	08ZR76	08ZR76 salmonella
20	339.5	17.5	361	16	08FJM9	08FJM9 escherichia
21	335	17.3	336	16	08FKD5	08FKD5 escherichia
22	329	16.9	337	2	09XBD3	09XBD3 amycolatops
23	328	16.9	335	16	08ZM60	08ZM60 salmonella
24	323	16.6	346	16	092MR5	092MR5 rhizobium m
25	310	16.0	369	16	08FB09	08FB09 escherichia
26	301.5	15.5	383	16	08FB02	08FB02 rhizobium l
27	268.5	13.8	332	16	08FCD6	08FCD6 escherichia
28	266	13.7	334	16	092YR7	092YR7 rhizobium m
29	261	13.4	345	16	08YB95	08YB95 bruceella me
30	261	13.4	345	16	08FX49	08FX49 bruceella su
31	255.5	13.2	332	16	08ZL83	08ZL83 salmonella
32	255	13.1	346	16	08U1X7	08U1X7 agrobacteri
33	254	13.1	332	16	08Z2C5	08Z2C5 salmonella
34	248	12.8	332	2	093Q64	093Q64 klebsiella
35	241	12.4	346	16	08U6Y0	08U6Y0 agrobacteri
36	240.5	12.4	358	16	08UB09	08UB09 agrobacteri
37	239	12.3	345	16	092MS8	092MS8 rhizobium m
38	228.5	11.8	333	16	098CD3	098CD3 rhizobium l
39	225	11.6	318	17	096YR3	096YR3 sulfolobus
40	166.5	8.6	190	2	08L314	08L314 corynebacte
41	105	5.4	2353	2	P71401	P71401 haemophilus
42	104.5	5.4	3275	16	08VKM3	08VKM3 mycobacteri
43	104.5	5.4	3300	16	006304	006304 mycobacteri
44	100.5	5.2	370	16	097FL4	097FL4 clostridium
45	100	5.1	845	5	08WQJ3	08WQJ3 plasmodium

ALIGNMENTS

RESULT 1
ID 062128 PRELIMINARY; PRT; 372 AA.
AC 062128; 062232;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE F36A2.3 protein.
GN F36A2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Felodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RL Leonard N.;
RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berke M.,
RA Bonfield J., Burton J., Connell M., Copeay T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighton J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RN SEQUENCE FROM N.A.
RL Leonard N.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z96047; CAB09417.1; -;
DR EMBL; Z81077; CAB09417.1; JOINED.

Dd		121	NHYGMAGNYAIRAMDQGLVGSMTNTSPLMAPTRAKREALCTNPBLSGANA--TNGDKFLL	179
Oy		183	DMATTVAATGVVELADCRGRKQIPDSTWADSKGNPSTQYVLHGGLPLPGIIEETGSY	242
			: : : : : : : : : : : : : : :	
Dd		180	DMATTAAVAVGVEIEIORRKG-APLPDGMAQDPDSGEVTNDAELFGSTGCLMGSGSELTSYG	238
Oy		243	KGTGLSMGVEFCGTLASSGFSFGKNVRM--GQSHKAADNGCCFYAIDQECFAPFAPRLQ	300
			: : : : : : : : : : : : : : : :	
Dd		239	KGYGGIAAVVDILSGMSGANSYTOYRKHTHAGAASADLIGOVFAIVPNCFAPNFEERMA	298
Oy		301	QFLDETRLKRKISEEKPVLVYGDEDERMNTEXSQKACGCVLYOEGOKALAELETATCDVQMF	360
Dd		299	DFNSRLRGATPTPSKRPVLLAGDEXEKGMADVDAAGIQYUENDUKTCANLAETLKTPPL	358
Oy		361	SY 362	
		:	:	
Dd		359	SF 360	
RESULT 4				
ID	O8IPU0	PRELIMINARY:	PRT:	361 AA.
Ac	O8IPU0:			
Dt	01-MAR-2003 (TREMBLrel. 23. Created)			
Dt	01-MAR-2003 (TREMBLrel. 23. Last sequence update)			
Dt	01-MAR-2003 (TREMBLrel. 23. Last annotation update)			
De	CG10512-PB.			
Gn	CG10512.			
Os	Drosophila melanogaster (Fruit fly).			
Oc	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Oc	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Oc	Ephydroidea; Drosophilidae; Drosophila.			
Rn	NCBI_TaxID=7227;			
Rx	[1]			
Rp	SEQUENCE FROM N.A.			
Rx	MEDLINE=20196006; PubMed=10731132;			
Ra	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
Ra	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
Ra	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
Ra	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
Ra	Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,			
Ra	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,			
Ra	Abiri J.F., Agbayani A., An H.J., Andrews-Ffiankko C., Baldwin D.,			
Ra	Ballew R.M., Basu A., Bakendale J., Bayraktaroglu I., Beasley E.M.,			
Ra	Beeson K.V., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
Ra	Borkova D., Botchan M.R., Bouck J., Brokslein P., Brotlier P.,			
Ra	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
Ra	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
Ra	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
Ra	Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
Ra	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
Ra	Foster C., Gabrieliann A.E., Garg N.S., Gelbart W.M., Glasser K.,			
Ra	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
Ra	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
Ra	Hoslin D., Houston K.A., Howland T.J., Wei M.H., Ibergam C.,			
Ra	Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
Ra	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
Ra	Lasko P., Lei Y., Levitsky A.A., Li J.Z., Li Z., Liang Y., Lin X.,			
Ra	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
Ra	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
Ra	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
Ra	Nelson D.R., Nelson K.A., Nixon R., Nusken D.R., Pacleb J.M.,			
Ra	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,			
Ra	Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,			
Ra	Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,			
Ra	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
Ra	Svirskas R., Teector C., Turner R., Venter E., Wang A.H., Wang X.,			
Ra	Wang Z.Y., Wassaman D.A., Weinscock G.M., Weltsbach J.,			
Ra	Williams S.M., Woodgerf, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,			
Ra	Yeh R.F., Zavertil J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
Ra	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,			
Ra	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,			
Rt	"The genome sequence of Drosophila melanogaster."			


```

OY 123 NHYGACOHYTKRIANAGMGSPFNTPSLMPPCSSSELTGNTPLSCVNSSEKTDGDFLL 182
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 206 NHYGAGWYAIRAMDQGLVGMSTNTSPLMAPTRAKPRAIGTNPLSIGANA--TNGDFFLL 264
OY 183 DMAITVALGKVELADRGKTOIPSTWGADSKGNPSTDTQVYLHGGLLPLPGIEEGSY 242
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 265 DMAITVAAYGKIELQRRKG--APLPDGAODPSGSEVTNDALGFGTGLMLPGSGSELTSGY 323
OY 243 KGTGLSMGELFCGIIILAGSSFGKNVRLM--GQSHKADNGCCFVAIDQECFAPGAPRLQ 300
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 324 KGYGLGAMVDILSGVMSGANYSQVRFKTHAGADSADLCQVFIAYDPNCFAPNFEERMA 383
OY 301 QFLDETRNLKPISEKRPVLPDPERMNTESQKAGGLVYQEGIKALEELATCDVOMF 360
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 384 DFNRLRGATPTDPSKPVLLAGDKERKGMADVDAAAGIOYLENDQKTCANLAEILKIKPL 443
OY 361 SY 362
    || : : : : :
Db 444 SF 445

```

RESULT 6

```

ID 096107 PRELIMINARY; PRT; 446 AA.
AC 096107;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE GH06154P.
GN CG10512.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY047551; AAK77283.1;
DR FLYBase; FBgn0037057; CG10512.
DR InterPro; IPR003767; Idh_2.
DR Pfam; PF02615; Idh_2; 1.
SQ SEQUENCE 446 AA; 47001 MW; 542A09F12BEE2207 CRC64;

```

```

Query Match 39.88; Score 773; DB 5; Length 446;
Best Local Similarity 44.88; Pred. No. 1.8e-56;
Matches 162; Conservativity 51; Mismatches 143; Indels 6; Gaps 5;

```

```

OY 4 SKDAPPEVVPKARREVFVKCMQVGRSPDHAGQLADLLDADLVGHSHGLNLTLYVD 63
    || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 87 SAAAPK-LVAVASRRPMICFRKAVKVPQAHAEQADLLVAADRGRHSHENLWEMTIN 145
OY 64 DVR-NGYKNGVPRVLKQKGTAMVDENLTGAVNGFCITDLATLKAKEFGVAVWVRKS 122
    || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 146 DLAINSTDGAAYPRILKETPATAMVDGLNGAVAYNGYCMDLATKKAITYGVAVCAKGS 205
OY 123 NHYGACOHYTKRIANAGMGSPFNTPSLMPPCSSSELTGNTPLSCVNSSEKTDGDFLL 182
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 206 NHYGAGWYAIRAMDQGLVGMSTNTSPLMAPTRAKPRAIGTNPLSIGANA--TNGDFFLL 264
OY 183 DMAITVALGKVELADRGKTOIPSTWGADSKGNPSTDTQVYLHGGLLPLPGIEEGSY 242
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 265 DMAITVAAYGKIELQRRKG--APLPDGAODPSGSEVTNDALGFGTGLMLPGSGSELTSGY 323
OY 243 KGTGLSMGELFCGIIILAGSSFGKNVRLM--GQSHKADNGCCFVAIDQECFAPGAPRLQ 300
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

```

```

Db 324 KGYGLGAMVDILSGVMSGANYSQVRFKTHAGADSADLCQVFIAYDPNCFAPNFEERMA 383
OY 301 QFLDETRNLKPISEKRPVLPDPERMNTESQKAGGLVYQEGIKALEELATCDVOMF 360
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 384 DFNRLRGATPTDPSKPVLLAGDKERKGMADVDAAAGIOYLENDQKTCANLAEILKIKPL 443
OY 361 SY 362
    || : : : : :
Db 444 SF 445

```

RESULT 7

```

ID 091771 PRELIMINARY; PRT; 385 AA.
AC 091771;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG18808 protein.
GN CG18808.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Neomates D., Celinker S.E., Gu Z., Guan P., Harris M.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.R., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sinden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003482; AAC92250.1;
DR FLYBase; FBgn0042131; CG18808.
DR InterPro; IPR003767; Idh_2.
DR Pfam; PF02615; Idh_2; 1.
SQ SEQUENCE 385 AA; 42081 MW; 65759AE3A998DE9F CRC64;

```

[illegible]

Query Match	22.2%;	Score 432;	DB 17;	Length 341;
Best Local Similarity	30.3%;	Pred. No. 5.4e-28;		
Matches 106;	Conservative 73;	Mismatches 153;	Indels 18;	Gaps 8

Query Match	21.0%	Score 408;	DB 16;	Length 356;
Best Local Similarity	31.7%	Pred. No.	6e-26;	
Matches 113; Conservative	63;	Mismatches	161;	Indels 20; Gaps 10
QY	12	VSEPKDAREFVVKQMOTGTS-PDHAQO-LADLLDDADLVGHSHGLNLRHLYVDYKNGV	69	
Dd	6	VSEKDIQRF---CRDLIKOSLPDEKEEVAETLTVADIRGYTSHGVQRAVATLKRMEDGI	62	
QY	70	KGNQVP-KYLKOKGTAWVDGENLLGAVVNGECTDLAIKLKEFGVANWVTKNSNHYGAC	138	
Dd	63	IERTKDISVYDFATALLDANNMGVOVAGVAKEMALIQKGOYTSPFVGVSNNHFCTA	122	
QY	129	QHRTKIANNAGVMGFSTPSPMPRCSSSELGLTNPLSCCVASNEKTGDSDLDMATT	188	
Dd	123	SFTYRMAAEKGFIIGIMETNASIPMVFEGAKEPSLDPNPISISIPAGEKKRPILIDMSTSN	182	
QY	189	VAVGWELADDCRKQTQIPSTWGADSKGNSTPYOVVLHGGLLPAGGIEETGSVKKHGLS	248	
Dd	183	VANQGKLTWAK-KKKEIPIEGMAITTDGKOTTADAME-GVLLPMPG-----KSGSLA	233	
QY	249	MMGELFCGLLAGSSFCKNVRLTGOSHKAADNQCFVALIDOCFAFG--FAPRLQOFDET	306	

Db 234 IIDILSLVLTGSLFGRIRPROYDDPYPOOLGLHFGALINENGEPELFFONLEKTEET 293
 QY 307 RNLPKPISEKPVLPVGPDE--RMNTEYSQKAGGLVYOGQIKALELTKCDVOMFSY 362
 Db 294 ISSEPSGPGQVYVMPGDMESQKRKEY--KENGIPLSGEIILNELKTEOKXGVLDLEH 348

RESULT 12

Q97LJ4 PRELIMINARY; PRT; 369 AA.
 AC 097LJ4;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE Malate dehydrogenase.
 GN CAC0566.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RC MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hatti J., Wolf Y.I.,
 RA Tatunov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum."
 RT J. Bacteriol. 183:4823-4838(2001).
 DR EMBL: AE007571; AK78545.1;
 DR InterPro: IPR003767; Idb_2.
 DR Pfam: PF02615; Idb_2; 1.
 KW Complete proteome.
 SO SEQUENCE 369 AA; 40989 MW; 589826988696C58 CRC64;

Query Match 21.0%; Score 407; DB 16; Length 369;
 Best Local Similarity 28.7%; Pred. No. 7.7e-26;

Matches 107; Conservative 76; Mismatches 164; Indels 26; Gaps 11;

QY 1 MNYSKDAPEFVVPKRDAREFVVKCMQVGTSPDHAGLADLLDADLVGHYSGNLNHI 60
 Db 8 VGSXKYEYEGELN-----LCLEVPORLGTTKEDSTYIGEVLLADLFGIESHGVRQLTL 61
 QY 61 YVDDVKNK-VKNGVCPVKLKQKGTAWVDENLGAVNGFCYDLAIKLAKRGVAVVT 119
 Db 62 YKGIIRGSIKVNHNHIVKESPVSVIDDGGGQIISKKAMERAIIDAKKKTGIGMAIV 121
 QY 120 KSNNHGACQHYTKKIANAGMVGMSFTNPSPLMPCRSSEIGLGNPLSCCVNSEKTGDS 179
 Db 122 KSNHFGIAGYRMAERKGLMGIAMNTFAIVLPTVGRGOMLGTFNISIAMPARP--P 179
 QY 180 FLDDMATTVTALCKVELADCRGKTQIPSTWAGSDKGNPSTDTQVYL-----HGGGLIPG 234
 Db 180 FLDDMATTVTTRKREYYS--KNNOKLYGAMDDGKKTNPDKVLANGTGGGLPLG 238
 QY 235 GIBET--GSYKGTGLSMNGELFCGLAGSSFGKNVRLMGQSHKAADNGQC--FYAIDQECF 291
 Db 239 GDEIFGSGHKGYSFAVEFTSLSGGLSDEIR--GNNN---TNCVCAFAFADVGLF 293
 QY 292 --APGAAPRLQQLDETNRNLKPISEKPVLPVGPDEPERMNTTEYSQKAGGLVYOGQIKALE 349
 Db 294 DDKNALTEKMSAVMOKIRNSKRAYGASRTYHGEKE--LEAYEDRMENGIPINENTYIEIK 352
 QY 350 ELATKCDVOMFSY 362
 Db 353 EICDIYVDIDINNT 365

RESULT 13
 050353 PRELIMINARY; PRT; 361 AA.

AC 050353;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Hypothetical 39.4 kDa protein.
 GN ORF-361.
 OS Lactobacillus helveticus.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1587;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC15009;
 RA Thompson K., McConville K.J., McReynolds C., Foley S.;
 RT "Complete sequence of plasmid pLH1 from Lactobacillus helveticus
 RT ATCC15009."
 RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AJ222725; CAA10971.1;
 DR InterPro: IPR003767; Idb_2.
 DR Pfam: PF02615; Idb_2; 1.
 DR PROSITE: PS00178; AA-TRNA-LIGASE_1; 1.
 KW Hypothetical protein; Plasmid.
 SO SEQUENCE 361 AA; 39426 MW; 10642F63B9C36C30 CRC64;

Query Match 19.4%; Score 377.5; DB 2; Length 361;
 Best Local Similarity 29.7%; Pred. No. 2.2e-23;

Matches 107; Conservative 65; Mismatches 161; Indels 27; Gaps 10;

QY 12 VSPRDAREFVVKCMQVGTSPDHAGLADLLDADLVGHYSGNLNHIYVDDVKNK-VK 70
 Db 3 ISSISEKEFEKVFARQFSKSDGALIDLVADIRKGISSHGIORLAWTSMIKKHIE 62
 QY 71 GNVCPVKLKQKGTAWVDENLGAVNGFCYDLAIKLAKRGVAVVTKSNHAGACOH 130
 Db 63 PQNKLVIKETPISLIDANKNGQIATAPAEKQIKTKYOKLSAVVIRNSHHFTAGT 122
 QY 131 YTKKIANAGMVGMSFTNPSPLMPCRSSEIGLGNPLSCCVNSEKTGDSFLDDMATTVYA 190
 Db 123 YARMAAKQGLIGIALNTRPLVVPNATEFLSNFAFPLPAEP--HPVFGATSIYS 180
 QY 191 LKVELADCRGKTQISTWAGSDKGNPSTDTQVYL-----GGGLIPLGIEE 238
 Db 181 SGRIQLL-AKNNOPIPGDMAVDNHRVMDQAQVYENNLAKVAETEKQPGGVLTIGGLQ 239
 QY 239 TGS-YKGTGLSMNGELFCGLAGSSFGKNVRLMGQSHKAADNGQCFVAIDQECFA--PGF 295
 Db 240 SNSNYGFGNSLVVELITGLAOGSISADINKGHH---DSQFLITINPELFGMDNL 295
 QY 296 APRLOQFLDETNRNLKPISEKPVLPVGPDEPERMNTTEYSQKAGGLVYOE--GOIKAL--EELA 352
 Db 296 KNSAEEMFRLRLKRLPGAE--IMIPGDRRYRYAENLKGQVYIDDKYVAELKTIQEBELA 354

RESULT 14

Q93004 PRELIMINARY; PRT; 334 AA.
 AC 093004;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
 DE Sma0265 protein.
 GN RAO141 OR Sma0265.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RC MEDLINE=21396509; PubMed=11481432;

RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 RA Gujal M., Hong A., Huizai L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yen K.-C., Davis R.W., Federspiel N.A., Long S.R.,
 RT "Nucleotide sequence and predicted functions of the entire
 RT *Strombolobium melioli* pSMA megaplasmid."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL: AE007208; AAK64799.1; -
 DR InterPro: IPR003767; Idb_2.
 DR Pfam: PF02615; Idb_2; 1.
 KW Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 347 AA; 34769 MW; 16E60916304D2F85 CRC64;

Query Match 19.1%; Score 371; DB 16; Length 334;
 Best Local Similarity 30.8%; Pred. No. 7e-23;
 Matches 101; Conservative 50; Mismatches 157; Indels 20; Gaps 6;

QY 14 PKDAREFVVKCMQVGTSPDHAGQLADLLDADLVGHSHGILNRLHIYVDVKNGV---- 69
 DB 5 PARLRNLSVALLEKRGVPADSRIOANLLLEALRGLEPSHGLQRLPLLSRLDKGLAMPT 64
 QY 70 -KGVGVKYLKQKGGTAVDGENLLGAVVGFCTDLAIKAKKEGVAVVTKNSNHGAC 128
 DB 65 TRNGGT----WRRASFSLVDGERGIGPVYMDAMKRVTRILKETGLATAIRNANMHGML 120
 QY 129 QHYTKKIANAGVMSFTNTSPLMFPCRSSEIGTNPISCCVNSEKTDSEFLDMATT 188
 DB 121 AYIEAARDELIGIVNSTSALVHPGGTALIGTNPVA--IGIPAAHGHPVLDLATSI 178
 QY 189 VALKVELADCRGKTQIPSTWGADSKGNPSTQVVLHGGLPLGIEETGSYKGTLS 248
 DB 179 VSMGKINNHAARG-LAIPGVAVDRDGRATIDPHA-----QAGATAIPFGDAKGYGLG 230
 QY 249 MGGLFCGILLAGSSFGKNVRLMGOSHKADNGCCFVAIDQCFAPGAPRLQQLDETRN 308
 DB 231 LAIELVAALAGSNLAPDVNCTLDDIHPANKGDLILIDPSAGA-GSIPALAAVYLDRLRL 289
 QY 309 LKPISEKPVLPDGPPERMNTEYSOKAG 336
 DB 290 SRPDPPTQPVAIPEGDARARRAAAKTG 317

RESULT 15
 O8NSK9 PRELIMINARY: PRT: 347 AA.
 AC O8NSK9.
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Malate/L-lactate dehydrogenases (EC 1.1.1.82).
 GN CG10661.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxId=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP005276; BAB98054.1; -
 DR InterPro: IPR003767; Idb_2.
 DR Pfam: PF02615; Idb_2; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 347 AA; 37134 MW; EDC18BFC9E9E1B55 CRC64;

Query Match 18.9%; Score 368; DB 16; Length 347;
 Best Local Similarity 31.0%; Pred. No. 1.3e-22;
 Matches 110; Conservative 53; Mismatches 164; Indels 28; Gaps 10;

QY 8 PEVVPKDAREFVVK--MOTVGTSPDHAGQLADLLDADLVGHSHGILNRLHIYVD 64

DB 2 PEVTV---NAQLFVLCTDILTKTGVPADADHLVGDLSLVQADLMCHPSHGVLRLPWYVR 58
 QY 65 VKNG-VKNGVPKVYLKQKGGTAVDGENLLGAVVGFCTDLAIKAKKEGVAVVTKNSN 123
 DB 59 LHSGMTTHAVVEVLNDIGAVLADGHNIGOVLAHDARKKRAVTRAMMFGIAYSVRSN 118
 QY 124 HYGACQHYTKKIANAGVMSFTNTSPLMFPCRSSEIGTNPISCCVNSEKTDSEFLD 183
 DB 119 HEGTAMYTTRKRAAGCVSILTNNASPMAPMGGEKIKITNPMSIAAPSET--ATVVD 176
 QY 184 MATTVVALGVELADCRGKTQIPSTWGADSKGNPSTQVVLHGGLPLGIEETGSYK 243
 DB 177 IANTAVARGKIYHAR-QTNMPLPETWALITSGAPTTDPAEAIN-GVLPMPAG-----HK 228
 QY 244 GTGLSMGELFCGILLAGSSFGKNVRLMGOSHKADNGCCFVAIDQCF--APGFAPRLQ 301
 DB 229 GYALSFMADVLSGVLTGSHSTKVHGPYDPPPGGAGHIFTALDVAAFRDPQDFDALS 288
 QY 302 FLDETRNLKPISEKPVLPDGPPE---RMNTEYSOKAGGLVYOBGQIKALEELA 352
 DB 289 LVGEVKSIPKAQNTFEITYPGESEDRARRKNS-----AHGISLPKXTMELQEL 338

Search completed: July 24, 2003, 21:14:52
 Job time : 112 secs

SHAP PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:09:49 ; Search time 42 Seconds

(without alignments)
838.041 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 1942

Sequence: 1 MNYSKDAPEFVSPKDAREF.....ALEELATKCDVMFSYKRLK 366

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	51.9	372	2	T20396 hypothetical prote
2	924	47.6	400	2	T18570 probable malate de
3	511	28.3	360	2	B71073 probable malate de
4	501	25.8	362	2	H75131 malate dehydrogena
5	465.5	24.0	339	2	S08981 malate dehydrogena
6	456.5	23.5	344	2	H64477 malate dehydrogena
7	407	21.0	341	2	H69027 malate dehydrogena
8	407	21.0	369	2	F96969 malate dehydrogena
9	390.5	20.1	337	2	E69852 malate dehydrogena
10	389.5	19.4	355	2	T20825 hypothetical prote
11	377.5	19.1	361	2	T46704 hypothetical prote
12	371	18.7	334	2	E95279 probable (imported
13	363.5	18.6	349	2	AB0568 ureidoglycolate de
14	360.5	18.6	349	2	C90701 ureidoglycolate de
15	360.5	18.6	349	2	F85551 malate dehydrogena
16	356.5	18.4	342	2	D64783 malate dehydrogena
17	331.5	17.1	361	2	G90738 probable dehydroge
18	331.5	17.1	361	2	A85589 probable dehydroge
19	326.5	16.8	361	2	A64817 malate dehydrogena
20	269.5	13.9	332	2	A65157 hypothetical prote
21	268	13.8	349	2	I38535 probable L-lactate
22	266	13.7	334	2	D95361 probable malate de
23	261	13.4	345	2	AD3635 malate dehydrogena
24	257.5	13.3	332	2	C64165 hypothetical prote
25	255	13.1	346	2	AF2596 malate dehydrogena
26	254	13.1	346	2	H97378 probable L-malate
27	254	13.1	332	2	AD0979 probable carboxyl
28	241	12.4	345	2	AH3131 malate dehydrogena
29	241	12.4	346	2	C98156 probable L-malate

30	240.5	12.4	358	2	A98332 malate dehydrogena
31	239	12.3	345	2	C95873 probable malate de
32	227.5	11.7	309	2	AB2951 malate dehydrogena
33	227.5	11.7	334	2	B83488 probable L-malate
34	104.5	5.4	3300	2	D70575 probable PPE prote
35	100.5	5.2	370	2	A97235 RCI1 repeats prote
36	98.5	5.1	1145	2	G87284 hypothetical prote
37	98	5.0	829	2	E87305 TonB-dependent rec
38	95.5	4.9	382	2	T49762 hypothetical prote
39	95	4.9	543	2	S46098 probable acid-CoA
40	94.5	4.9	382	2	I39780 subtilisin (EC 3.4
41	94	4.8	512	2	A87270 hypothetical prote
42	94	4.8	2154	2	F83068 hypothetical prote
43	93.5	4.8	540	2	F96841 hypothetical prote
44	93	4.8	487	2	T49424 hypothetical prote
45	93	4.8	1101	2	T21062 hypothetical prote

ALIGNMENTS

```

RESULT 1
T20396
hypothetical protein F36A2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T20396; T21836
R:Lennard, N.
submitted to the EMBL Data Library, June 1997
A:Reference number: Z19266
A:Accession: T20396
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-372 <MWL>
A:Cross-references: EMBL:Z66047; PIDN:CAB09417.1; GSPDB:GN00019; CESP:F36A2.3
A:Experimental source: clone DY3
R:Lennard, N.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19476
A:Accession: T21836
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-372 <MWL>
A:Cross-references: EMBL:Z81077; PIDN:CAB03073.1; GSPDB:GN00019; CESP:F36A2.3
A:Experimental source: clone F36A2
C:Gene: CESP:F36A2.3
A:Map position: 1
A:Intons: 21/3; 53/2; 79/1; 114/3; 147/3; 198/3; 249/1; 284/3; 353/3
C:Superfamily: malate dehydrogenase yIbC

Query Match          51.9%; Score 1008; DB 2; Length 372;
Best Local Similarity 54.6%; Pred. No. 1.2e-72;
Matches 194; Conservative 48; Mismatches 111; Indels 2; Gaps 2;

QY      6 DAPFVSPKDAREFVSKMOTVGTSPDAGLADLLDADLVGHYSHGLNHTYVD- 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     11 ETDEIVTSKELDSFVLECLAKAGCTGDAQDLAEFLCSDYRGHSHGIRLHYVDL 70
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     65 VNGVYGVNPKVYLKOKGCTAVNDENILGAVVNGCTDLAKLKEFVAVVYKNSNH 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     71 MKKSTAVVGTPOVLKSKGSTAVDGNLLGPVYVNGCMQLAIVKAKKEFGVYVCRNSH 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     125 YVACOHTYTKINAGVGSFTNPSLMPFPCRSSEIGTNPVLSGVNSKTSGLDMD 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     131 FGIAGVYADFACRNGIYGVAFNTSCVPTGSREKSLGSDNP1-CMAAPGMGDSFFLDM 189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     189 ATTGTVALGVELADRGKQIPTSGAGSKGNSPTDQVYVHLGGILLPGIEETGSYKG 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     190 ASITVAYGKTEVYDKRGETIYIGSWGADNGDETHNPKVYLDGGGLQPGSGSEITGKYG 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     245 TGLSMNGELFCGIIAGSSFGKKNVRLMGOSHRKADNGOCFVAIDOCFAFGFAPRLQOFLD 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

QY 68 GVGKNGVPKYLKQGGTAWVDGENILGAVNGNCTDLAIKLAKERGVAMVTKNSNHYA 127
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 71 GINLHPRIKIRREBPATALLDDEGFGVGVYKAMKLAIEKARTGIGIVARNNSHYCI 130
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 128 CQHTTKIANAGVAMGSEFNTSPLMFPCRSSEITGLTNPLSCVNSEKTDGSLDMATT 187
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 131 AGYALMAAEKMGISMTNSRPLVAPGVERILGTNPILAPTK--GKPFLLDMATS 188
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 188 TVALGVELADCRKTOIPSTWAGDSKGNPSTQVYVHGGILLPLGIEE--TGSYKGTG 246
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 189 VVPGIKLEVRKRKE--ELPEGWAINSGELTRVSEVFNGLPLGSGFGLGCHKGYG 247
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 247 LSMAGLEFCGILAGSSFGKNVRLWGOSHKADNGOCFAVADQECFAP--GFAPRLOQFLD 304
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 248 LSLAVDILSGTISGTSKSHYK--NTNKNNSNVCHFEPAITIEHTPLPEEKGKNSMIN 305
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 305 ETRNLKPISEKPYLVPGDPERMNTESQKAGGLVYOGQIKALEEL 351
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 306 EIKNSRKHPDERIRIWHGKGFLLMETRLKLGIDYK---KVIDEL 348
   | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

RESULT 5

```

S08981
malate dehydrogenase (EC 1.1.1.37) - Methanothermus fervidus
C:Species: Methanothermus fervidus
C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 13-Sep-1998
A:Accession: S08981; S08689
R:Honka, E.; Fabry, S.; Niermann, T.; Palm, P.; Hensel, R.
Eur. J. Biochem. 188, 623-632, 1990
A>Title: Properties and primary structure of the L-malate dehydrogenase from the extreme
A:Reference number: S08981; MUID:90235834; PMID:2110059
A:Accession: S08981
A:Molecule type: DNA
A:Residues: 1-339 <HON>
A:Cross-references: EMBL:X51840
R:Honka, E.; Fabry, S.; Niermann, T.; Palm, P.; Hensel, R.
submitted to the EMBL Data Library, February 1990
A:Reference number: S08689
A:Accession: S08689
A:Molecule type: DNA
A:Residues: 1-2-339 <HO2>
A:Cross-references: EMBL:X51714
C:Genetics:
A:Start codon: TTG
C:Superfamily: malate dehydrogenase yJbc
C:Keywords: oxidoreductase, tricarboxylic acid cycle

```

```

Query Match 24.0%; Score 465.5; DB 2; Length 339;
Best Local Similarity 33.0%; Pred. No. 1.8e-29;
Matches 116; Conservative 67; Mismatches 150; Indels 19; Gaps 8;

```

```

QY 11 VVSPKDAREFVVKMQVGTSPDHAGQLADLLDADLVGHYSHGLNRLHIYVDYKNG-V 69
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 2 IISPEERSLIIKILNLGVSEEHAKITADIVADLKGFTSHGGRPOYVEGKILGTI 61
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 70 KGNVPPKVLKQGGTAWVDGENILGAVNGNCTDLAIKLAKERGVAMVTKNSNHYACQ 129
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 62 KTSNIEIEKETSVALINGNHLIGQVYVAKMKLAIEKARTGIGIVARNNSHYCIAG 121
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 130 HYTKKIANAGVAMGSEFNTSPLMFPCRSSEITGLTNPLSCVNSEKTDGSLDMATT 189
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 122 YVSDMAKNDMIGITMTTEPAVAPLGGKIPVLTNPFAISIPSNFY--YVAVDMSTAAV 179
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 190 ALGVELADCRKTOIPSTWAGDSKGNPSTQVYVHGGILLPLGIEETSYKGTGLSM 249
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 180 ARGTL-LEARKNKIEPIGAVNDKNGNPTTNPALN--GSILPFGG-----HKGVALCF 231
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 250 MGELEFCGILAGSSFGKNVRLWGOSHKADNGOCFAVADQECF--APGAPRLOQFLDET 307
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 232 MIELAGPLVAKEGSKYKGTVPDSQOMCTKGLDLAIDPSKFYDLEEKRVNDEVAKIK 291
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 308 NLKPISEKPYLVPGDPERMNTESQKAGGLVYOGQIKALEELATKCDVGM 359
   | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

```

Db 292 -----STGKDVILIPDREERNIKRKE--GIELDKKIVLEKLEIADLIEL 337
   | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

RESULT 6

```

H64477
malate dehydrogenase (EC 1.1.1.37) - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
A:Accession: H64477
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
  , Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glöck
  , rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
  Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
  A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
  A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: H64477
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-344 <BUL>
A:Cross-references: GB:U67583; GB:L77117; NID:g2826409; PIDN:AA899436.1; PID:g1592075
C:Genetics:
A:Map position: FOR1395033-1396067
C:Superfamily: malate dehydrogenase yJbc
C:Keywords: oxidoreductase

```

```

Query Match 23.5%; Score 456.5; DB 2; Length 344;
Best Local Similarity 32.9%; Pred. No. 9.5e-29;
Matches 117; Conservative 64; Mismatches 156; Indels 19; Gaps 10;

```

```

QY 11 VVSPKDAREFVVKMQVGTSPDHAGQLADLLDADLVGHYSHGLNRLHIYVDYKNGV 70
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 2 ILKPENEKRLIIVLAKRGVPEEDAKITADVFDVADLKGFTSHGGRPOYITALKLG-N 60
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 71 GNGVP--KVLKQGGTAWVDGENILGAVNGNCTDLAIKLAKERGVAMVTKNSNHYAC 128
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 61 INPKPDIKIVKESPAVIDGDLGIGYVKKAMELAIKKANGVAVARNANHFGLA 120
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 129 QHYTKKIANAGVAMGSEFNTSPLMFPCRSSEITGLTNPLSCVNSEKTDGSLDMATT 188
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 121 GYVSLANNQDMIGITMTTEPAVAPLGGKIPVLTNPFAISIPSNFY--YVAVDMSTAAV 178
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 189 VALGVELADCRKTOIPSTWAGDSKGNPSTQVYVHGGILLPLGIEETSYKGTGLS 248
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 179 IARGKI-LEALRKTKIPIEGCAVNDKNGNPTTNPALN--GCILPFGG-----PKGYLA 230
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 249 MGELEFCGILAGSSFGKNVRLWGOSHKADNGOCFAVADQECF--APGAPRLOQFLDET 306
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 231 LAIEMLSAI--GGAEGVTKVKGSTANPEEECTKGLDLEIAINPEFFMGKEEFKRVDELDEI 289
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 307 NLKPISEKPYLVPGDPERMNTESQKAGGLVYOGQIKALEELATKCDVQMSY 362
   | : : : : : | : : : : | : : : : | : : : : | : : : : |
Db 290 KNSP-AGFELLIPGEIERNK--MKRKDFEIDKNLYNOLKEICNELGNTIEDY 342
   | : : : : : | : : : : | : : : : | : : : : | : : : : |

```

RESULT 7

```

H69027
malate dehydrogenase - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
A:Accession: H69027
R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T
  ; Qiu, D.; Spadafora, R.; Vitacore, R.; Wang, Y.; Mierzbowski, J.; Gibson, R.; Jivan
  , i, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
  J. Bacteriol. 179, 7135-7155, 1997
A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
  A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: H69027
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-341 <WH>
A:Cross-references: GB:AE000888; GB:AE000666; NID:g2623304; PIDN:AA885694.1; PID:g262
  A:Experimental source: strain Delta H

```

C:Genetics:
A:Gene: MTH1205
C:Superfamily: malate dehydrogenase yJbc

Query Match 21.0%; Score 407; DB 2; Length 341;

Best Local Similarity 29.6%; Pred. No. 8.3e-25;
Matches 104; Conservative 77; Mismatches 150; Indels 20; Gaps 8;

QY 12 VSPRDAREFVVKCMQTVGTSPPDHAGOLADLLDADLVGHSHGLNRLHIYDDVKNKNG-VK 70
DB 3 ISPEBEKIIKELITANNVPESSDIADVTLDADLKGFSSHGKGRPOVYDGRHGTIR 62
QY 71 GNGVPKYLKQKGTAWVDGNLLGAVVGNFCTDLAIKLAKEGVAVVTKNSNHYGACQH 130
DB 63 ADGDITIERETESTALINGHIFGHVAVYRAMELAIEKANTGVLGVHDSNHFVAGY 122
QY 131 YTKKINAGVMSFTTSPMLPCRSSEIGCTNPSCCVNSKTKDSFLDMATTVA 190
DB 123 YSDMAVANDMIGVIAITEPAVPIGGRKPIIGTNPVIAIIPSNRYVS--VDMATISASA 180
QY 191 LGHVELADCRGKQIPISTWGDASKGNPSTDTQVVLHGGGLPLGIEETGSGYKGLISM 250
DB 181 RGLLEAKRKGES-IPENVALDAEGKRTDPEMAK-GSTILPFG-----HKGVALSEFM 232
QY 251 GELFCGILAGSSFGKKNVRLMGOSHKADNGOCFVAIDQCEPA--PGFAPRIQOFLDETRN 308
DB 233 IETILGVGAAGTAATGTANPEEMCKTGDLMMAIDPSKMPVDFEFRAQVDEFIEE--- 289
QY 309 LKPISEKPYLVPGDPERMNTYESOKAGLVYQEGOKALELATKCDVOM 359
DB 290 ---VKSSGDVILIPGDIIESMNIK-RRREGIELEDEKLERILGARELDINL 336

RESULT 8

F96969
malate dehydrogenase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: F96969
R:Rolling, J.; Bretton, G.; Omelchenko, M.V.; Markarov, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Reference sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*
A:Accession: F96969
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-369 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78545.1; PID:q15023433; GSFDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0566
C:Superfamily: malate dehydrogenase yJbc

Query Match 21.0%; Score 407; DB 2; Length 369;

Best Local Similarity 28.7%; Pred. No. 9.2e-25;
Matches 107; Conservative 76; Mismatches 164; Indels 26; Gaps 11;

QY 1 MNTSKDAPEFVVKCMQTVGTSPPDHAGOLADLLDADLVGHSHGLNRLHI 60
DB 8 VGSSTKRYEGILN-----LCEVFORLGTTKEDSTITGEVLLADLFGIESHGVOURLTL 61
QY 61 YDDVKNKNG-VKNGVPKYLKQKGTAWVDGNLLGAVVGNFCTDLAIKLAKEGVAVVTV 119
DB 62 YKGIIRNGSIKVNKNHNIYVSPVYIDGDGMOGQIISKAMERAIIDAKKTKGIGAIY 121
QY 120 KSNNHGACOHYTKKIANAGVMSFTTSPMLPCRSSEIGCTNPSCCVNSKTKDS 179
DB 122 KSNNHGACOHYTKKIANAGVMSFTTSPMLPCRSSEIGCTNPSCCVNSKTKDS 179
QY 180 FLDDMTATTVVALKVELADCRGKQIPISTWGDASKGNPSTDTQVVL-----HGGGLPLG 234
DB 180 FLDDMTATTVVALKVELADCRGKQIPISTWGDASKGNPSTDTQVVL-----HGGGLPLG 234

QY 235 GIEET-GSYKGTISNMGELECGILAGSSFGKKNVRLMGOSHKADNGOC--FVAIDQCEP 291
DB 239 GDEIRGKHGKYGKISFVKEFTSILSGLSDEIR--GNNN---TNGVCHAFADIDYGLF 293
QY 292 --APGAPRIQOFLDETRNPKPISEKPYLVPGDPERMNTYESOKAGLVYQEGOKALE 349
DB 294 DKNALIEKRMASAVMCKIRNSKKAAYGASRIYTHGEKE-LEAYEDRMENGIPINENTIVEIK 352
QY 350 ELATKCDVOMFSY 362
DB 353 EICDYVDIDINNY 365

RESULT 9

F69852
malate dehydrogenase homolog yJmc - Bacillus subtilis

C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: E69852
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fuma, S.; Galluzzi, A.; Gal lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl A:Authors: Schleich, S.; Schroeter, R.; Scofield, P.; Sekiguchi, J.; Sekowska, A.; Se akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchuya T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69852
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
A:Cross-references: GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CA813089.1; PID:g26335
A:Experimental source: strain 168
C:Genetics:
A:Gene: yJmc
C:Superfamily: malate dehydrogenase yJbc

Query Match 20.1%; Score 390.5; DB 2; Length 337;

Best Local Similarity 31.6%; Pred. No. 1.7e-23;
Matches 104; Conservative 66; Mismatches 142; Indels 15; Gaps 6;

QY 12 VSPRDAREFVVKCMQTVGTSPPDHAGOLADLLDADLVGHSHGLNRLHIYDDVKNKNGVXG 71
DB 6 IAAEKEKELVWQKLDAGLNERDAEKVADVADLARNHSHGVLTENHYVNLGAGIN 65
QY 72 NGVPKYLKQKGTAWVDGNLLGAVVGNFCTDLAIKLAKEGVAVVTKNSNHYGACQH 130
DB 66 PGAQAPKEKGTGPTVGLDGDGFGHNCMDAMAHADMAKKKGVAVTAVNSHCCALSY 125
QY 131 YTKKINAGVMSFTTSPMLPCRSSEIGCTNPSCCVNSKTKDSFLDMATTVA 190
DB 126 FVQKADDEKILGAMHTDTSIVPFGGRPIIDTNPFAVGPV-KKKRPILDMATSKVA 184
QY 191 LGHVELADCRGKQIPISTWGDASKGNPSTDTQVVLHGGGLPLGIEETGSGYKGLISM 250
DB 185 FKKILIDAREGK-EIEBGMDVNDNGEAVYDPOKVV-----SLSTGGPGRYGLSLV 234
QY 251 GELFCGILAGSSFGKKNV-RLMGOSHKADNGOCFVAIDQCEPA--PGFAPRIQOFLDETR 307
DB 235 VDFSGILAGAAAGPHIAKKNYINLDQKRKIGHVCAINISFDTMDTFLEOMAMIDEQ 294
QY 308 NLKPISEKPYLVPGDPERMNTYESOKAG 336
DB 295 QSPPAVGFERYVYVGEIEOLHEERKNKG 323

T20825

[illegible]

E955279

QY 309 LKPISEKPVLPDPERMNTIEYSOKAG 336
.. . . .

RESULT 13

AB0568

ureidoglycolate dehydrogenase (EC 1.1.1.154) [imported] - Salmonella enterica subsp. ent

C:Species: Salmonella enterica subsp. enterica serovar Typh

A:Note: This species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AB0568

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gea, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AB0568

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05012.1; PID:q16501795; GSPDB:GN00176

C:Genetics:

A:Gene: allid

C:Superfamily: malate dehydrogenase ylbC

C:Keywords: oxidoreductase

Query Match 18.7%; Score 363.5; DB 2; Length 349;

Best Local Similarity 30.7%; Pred. No. 2.5e-21;

Matches 102; Conservative 55; Mismatches 154; Indels 21; Gaps 10;

QY 12 VSPDAREFVVKCQQTGTSPPDHAGQLADLLDADLVGHYSHGILNRLHIYVDYKNGVKG 71

Db 3 ISRETLHLQLENNKLCQAGLKRHAATYAEVLVADANGISHGAVREYAEIRISKG-GT 61

QY 72 NGVPKVLKQKGG--TAWVDENLIGAVVNGCTDLAIKLAKEGVAVVTKNSHYGACQ 129

Db 62 NRETEPFLTEGTPCSTALIHADNAGQVAAKMGHAIKTAQNGVAVVYISRMGSGAIS 121

QY 130 HYTKIANAGVNGSFTNTSPDLMPCCRSSEIGLTNPISCCVNSEKTDGDFL-LDMAATT 188

Db 122 YFVQOARAGLIGISICQSDPMVVPFGGADITYCTNPLAFAPSE--GDDITFTDMATT 179

QY 189 VALGKVELADCRGKT-QIPSTWGDASKGNPSTDPQVVLHGGGLPLGIEETGSGYKGTG 247

Db 180 QAWGKV--LDARSNMSIPESMAVDKNGAPHTDFPFAV--NALLPAG-----PKGYGL 228

QY 248 SMGELFCGILAGSGFGKNV-RLMGQSHKADNGCCEVAIDQECFAPG--FAPRLQDFID 304

Db 229 MMMDIILSGILGLPFGROYSSMYEDLHAGRNQLGHLVYINPAFFSSCELFKRIHSQTMQ 288

QY 305 ETRNLKPLISEKPYLVPGDPERMATEYSQKAG 336

Db 289 ELNSVKKPAPGFKQYVYPGDDIKOKNADMMNG 320

RESULT 14

C90701

ureidoglycolate dehydrogenase [imported] - Escherichia coli (strain O157:H7, substrain H

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: C90701

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaswara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinaagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A9629; MUID:21156231; PMID:11258796

A:Accession: C90701

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <NAV>

A:Cross-references: GB:BA000007; PIDN:BA834002.1; PID:q13360037; GSPDB:GN00154

C:Genetics:

A:Experimental source: strain O157:H7, substrain RMD 0509952

A:Gene: ECs0579

C:Superfamily: malate dehydrogenase ylbC

Query Match 18.6%; Score 360.5; DB 2; Length 349;

Best Local Similarity 31.8%; Pred. No. 4.3e-21;

Matches 102; Conservative 49; Mismatches 149; Indels 21; Gaps 10;

QY 12 VSPDAREFVVKCQQTGTSPPDHAGQLADLLDADLVGHYSHGILNRLHIYVDYKNGVKG 71

Db 3 ISRETLHLQLENNKLCQAGLKRHAATYAEVLVADANGISHGAVREYAEIRISKG-GT 61

QY 72 NGVPKVLKQKGG--TAWVDENLIGAVVNGCTDLAIKLAKEGVAVVTKNSHYGACQ 129

Db 62 NRETEPFLTEGTPCSTALIHADNAGQVAAKMGHAIKTAQNGVAVVYISRMGSGAIS 121

QY 130 HYTKIANAGVNGSFTNTSPDLMPCCRSSEIGLTNPISCCVNSEKTDGDFL-LDMAATT 188

Db 122 YFVQOARAGLIGISICQSDPMVVPFGGADITYCTNPLAFAPSE--GDDITFTDMATT 179

QY 189 VALGKVELADCRGKT-QIPSTWGDASKGNPSTDPQVVLHGGGLPLGIEETGSGYKGTG 247

Db 180 QAWGKV--LDARSNMSIPDTWAVDKNGAPHTDFPFAV--NALLPAG-----PKGYGL 228

QY 248 SMGELFCGILAGSGFGKNV-RLMGQSHKADNGCCEVAIDQECFAPG--FAPRLQDFID 304

Db 229 MMMDIILSGILGLPFGROYSSMYEDLHAGRNQLGHLVYINPAFFSSCELFKRIHSQTMR 288

QY 305 ETRNLKPLISEKPYLVPGDPE 325

Db 289 ELNATTPAPGFKQYVYPGDD 309

RESULT 15

F85551

malate dehydrogenase (EC 1.1.1.37) probable [imported] - Escherichia coli (strain O15

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: F85551

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

11ler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85551

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <SHO>

A:Cross-references: GB:AE005174; NID:q12513412; PIDN:AA654874.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ylbC

C:Superfamily: malate dehydrogenase ylbC

C:Keywords: oxidoreductase

Query Match 18.6%; Score 360.5; DB 2; Length 349;

Best Local Similarity 31.8%; Pred. No. 4.3e-21;

Matches 102; Conservative 49; Mismatches 149; Indels 21; Gaps 10;

QY 12 VSPDAREFVVKCQQTGTSPPDHAGQLADLLDADLVGHYSHGILNRLHIYVDYKNGVKG 71

Db 3 ISRETLHLQLENNKLCQAGLKRHAATYAEVLVADANGISHGAVREYAEIRISKG-GT 61

QY 72 NGVPKVLKQKGG--TAWVDENLIGAVVNGCTDLAIKLAKEGVAVVTKNSHYGACQ 129

Db 62 NRETEPFLTEGTPCSTALIHADNAGQVAAKMGHAIKTAQNGVAVVYISRMGSGAIS 121

QY 130 HYTKIANAGVNGSFTNTSPDLMPCCRSSEIGLTNPISCCVNSEKTDGDFL-LDMAATT 188

Db 122 YFVQOARAGLIGISICQSDPMVVPFGGADITYCTNPLAFAPSE--GDDITFTDMATT 179

QY 189 VALGKVELADCRGKT-QIPSTWGDASKGNPSTDPQVVLHGGGLPLGIEETGSGYKGTG 247

Db 180 QAWGKV--LDARSNMSIPDTWAVDKNGAPHTDFPFAV--NALLPAG-----PKGYGL 228

Fri Jul 25 10:38:38 2003

us-10-060-848-3_1.rpr

Page 7

[illegible]

Search completed: July 24, 2003, 21:15:29
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:02:50 ; Search time 23 Seconds

748.338 Million cell updates/sec

```

Title:      US-10-060-848-3
Perfect score: 1942
Sequence:   1 MNYSKDAPEFVNSPKDAREF.....ALEELATKCDVQMSTYRLK 366

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
------------------	---------------	----

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	511	26.3	360	1	MDH_PYRHO	059028 pyrococcus
2	501	25.8	362	1	MDH_PYRAB	09v005 pyrococcus
3	465.5	24.0	339	1	MDH_METFE	P16142 methanobacter
4	456.5	23.5	344	1	COMC_METTJA	058820 methanococcus
5	407	21.0	341	1	COMC_METMTH	027273 methanobacillus
6	390.5	20.1	337	1	YJMG_BACSU	034736 bacillus subtilis
7	360.5	18.6	349	1	ALDD_ECO57	P58408 escherichia coli
8	356.5	18.4	349	1	ALDD_ECOLI	P58408 escherichia coli
9	331.5	17.1	361	1	YBIC_ECO57	P75555 escherichia coli
10	326.5	16.8	361	1	YBIC_ECOLI	P30178 escherichia coli
11	269.5	13.9	332	1	YIAK_ECOLI	P37672 escherichia coli
12	268	13.8	349	1	LDH_ALCEU	007251 alcalligenes
13	258	13.3	332	1	YC56_PASMU	09C115 pasteurella
14	257.5	13.3	332	1	YIAK_HAEIN	P44995 haemophilus
15	227.5	11.7	334	1	YC52_PSEAE	091492 pseudomonas
16	99	5.1	940	1	GBR2_RAT	088871 rattus norvegicus
17	97	5.0	379	1	RECA_STRTR	09ev77 streptococcus
18	97	5.0	655	1	ACDY_BOVIN	P48818 bos taurus
19	96.5	5.0	383	1	RECA_STRMU	P27624 streptococcus
20	96.5	5.0	860	1	VG12_BFB03	037893 bacteriophages
21	96	4.9	378	1	RECA_STRP3	08K5K0 streptococcus
22	92.5	4.8	543	1	FAT2_YEAST	P38137 streptococcus
23	92.5	4.8	363	1	EFUO_GYNST	063930 glycomicrobium
24	92.5	4.8	1017	1	KCRH_HUMAN	09u050 homo sapiens
25	92	4.7	941	1	GBR2_HUMAN	09u050 homo sapiens
26	92	4.7	1156	1	PHR3_SOYBN	075893 homo sapiens
27	91.5	4.7	916	1	GYRA_NEMGO	P42499 glycine max
28	91	4.7	378	1	RECA_STRP8	P48371 neisseria meningitidis
29	91	4.7	378	1	RECA_STRP1	08n230 streptococcus
30	90	4.6	494	1	AMR4_DROPS	059942 streptococcus
31	90	4.6	900	1	IF2_MYCUM	018552 drosophila
32	89.5	4.6	739	1	PUR1_LACIC	P71613 mycobacterium
33	88.5	4.6	424	1	GATD_METAC	09ZB06 lactococcus
						08U766 methanosaeta

34	87.5	4.5	333	1	YR55_IRV6	P18309	chlo_iride
35	87.5	4.5	575	1	FLA2_CAM6E	P22251	campylobact
36	87.5	4.5	728	1	SP61_P6A	Q43075	pium_sativ
37	87.5	4.5	997	1	YR42_CAE6L	Q20256	caenorhabdi
38	87.5	4.5	1023	1	RT11_ACT6PL	P55128	actinobacti
39	87	4.5	534	1	PROA_VIBAL	P16588	vibrio_algi
40	87	4.5	536	1	SPM1_MAG6R	P58371	magnaporthe
41	86.5	4.5	381	1	RECA_ST6PA	O85502	streptococc
42	86.5	4.5	387	1	YF20_MET6A	O58915	methanococc
43	86.5	4.5	394	1	EF9U_BOR6U	P25062	borrelia_bu
44	86.5	4.5	575	1	FLB2_CAM6E	P22252	campylobact
45	86.5	4.5	978	1	PM4_CHL6N	Q92895	chlamiidia p

ALIGNMENTS

```

RESULT 1
MDH_PYRHO          STANDARD;      PRI;       360 AA.
ID MDH_PYRHO
OS09028;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
MDH OR PH1277.
OS Pyrococcus horikoshii.
OC Archaeae; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_Taxid=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kwarababaya Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamita M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nekamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuwa H., Kikuchi H.;
RT "complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no way mo
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to licensel@isb-sib.ch).
CC -----
DR DR EMBL: AF000005, BAA30380.1; -.
DR PIR: B71073; B71073.
DR InterPro: IPR003767; Idh_2.
DR Pfam: PF02615; Idh_2.1.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
SQ SEQUENCE 360 AA; 39751 MW; 6E9D8B16ECDDEBE6F CRC64;
-----
Query Match 26.3%; Score 511; DB 1; Length 360;
Best Local Similarity 36.9%; Pred. No. 7.7e-35;
Matches 128; Conservative 66; Mismatches 139; Indels 14; Gaps 8
QY 10 EVVSPEDAR-EVVVVCMTGVTSPPHAQLDLIDLALDVGHSGNLRLHIYVD-VKN 67
DB 11 VIVNPDRLRFSLFIVITKTLGVRBEDDAIVADINVLMDLRKVESHGQRLLKRYVDGIIS 70
YY 68 GKAGNCVPVYLKOGGTIANVDGENTLGAVGNFCTDIAIKLAKEFGVAMVVTNSNHYA 127

```

Db 71 GVALHFKIRIRVIREGSPSYALIDDEDEGIGVYGRSMKLIKKRAKDPDGIIVTARNSNHYGI 130
 QY 128 CQHYYTKKINAGVNGVSPFTNPSLPIHPCCSSBIGTGTPMLSCVNSETKGSEFLDMATT 187
 Db 131 AGYIYALMAAEEGIGISTMTNSRPLVAPGIEIRILLGTNPMLAAPTK--DKPFLDMATS 188
 QY 188 TVALGKVELADCGKQKPISTPGADSKGNPSTDPVVLHGGLPLGITEE-TGSYKGTG 246
 Db 189 VVDIGKLEYRRRGK-DIEPGMAINREGNITTKVEEYVNGALLPLGGFGLLGCHNGYG 247
 QY 247 LSMAGELFGCIIAGSSFGKRVNRLMGOSHKADNGCCFYALDQECFAP--GFAPLQGLD 304
 Db 248 LSLMVDILSGILSGTWSKYRK-VNTESGSNVCHFFVVDIEHPIPLEEKERKISOMIE 305
 QY 305 ETENLKPISPEKPVLPDGPERRMNNYESOKAGLVLYGQIKALEEL 351
 Db 306 EIKSSKRHPFERIWHGEKGLVTHETRLKLDIPYR---KVFEEI 348

ID	MDH_PYRAB	STANDARD:	PRT:	362 AA.
AC	Q9Y0D5			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Malate dehydrogenase (EC 1.1.1.37).			
GN	MDH OR PYRAB08550 OR PAB1791.			
OS	Pyrococcus abyssi.			
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;			
OC	Pyrococcus.			
OX	NCBI_TaxID=29292;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GES / Orsay;			
RX	PubMed12622808;			
RA	Cohen G.N., Barbe V., Flament D., Galperin M., Helling R., Lecompte O.,			
RA	Poch O., Prieur D., Quenellou J., Ripp R., Thierry J.-C.,			
RA	Van der oost J., Welzenbach J., Zivanovic Y., Forterre P.;			
RT	"An integrated analysis of the genome of the hyperthermophilic			
RT	archaeon Pyrococcus abyssi."			
RL	Mol. Microbiol. 47:1495-1512(2003).			
CC	-1. CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.			
CC	-1. SUBUNIT: Homodimer (By similarity).			
CC	-1. SUBCELLULAR LOCATION: Cytoplasmic (Probable).			
CC	-1. SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ ,			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL, AJ248285; CABA9769.1; -			
DR	PIR: H75131; H75131.			
DR	InterPro: IPR003767; Idb_L2.			
DR	Pfam: PF02615; Idb_L2: 1.			
KW	Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.			
SQ	SEQUENCE 362 AA; 39907 MW; 1E15C2E9E6BA5012 CRC64;			
Query Match	25.8%;	Score 501;	DB 1;	Length 362;
Best Local Similarity	36.0%;	Pred. No. 5.2e-34;		
Matches 125;	Conservative 137;	Mismatches 14;	Gaps 8;	
QY	10 FVVSFKDAR-EEVVKCMQFTGTSPPHAQLDLDLADLVGHYSGILRLHYVDV-KN 67			
DB	11 YVRIPKDELFEFVVYRLTKLGTPEEDAKIYADNLIMDLDRGLSGVGQRLKRYVDGIIISG 70			
OY	68 GVKNGVGVKVLKQKGTAWVDGEMLLGAVGNFCTDIAIKLAKGVAVMVYTKNSNHYGA 127			
DB	71 GINAPRPRIKIREGVSVALDGDGEGGVGVKAKMLITFKARKGTGIVAAVRNSNNHGT 130			

OY		128	COHETKINACVAGSFPNTPLAFPCRRSSIGTIPMLSCVNSKRTGSEFLDMATT	187
Dd		131	AGYTALMAAEECHMIGISMTNSRPLVAPPGCVBRILGTNRIMAAPTK--GKPLFMDATS	188
OY		188	TVALGVELADCRKTOIPSTWGADSKGNPSTDQVYLHGGLLPLGGIEB-TGSYKGTC	246
Dd		189	VVPFIGLEYVRKGE-EIPEGMAINSKEITRSVEVEFGSLLPLGFGLLGHHGYC	247
OY		247	LSMGGLPGGILLAGSSFCKNVAYLMQSHKADNGCCFAIDQBEFAR--GPAPRLQGFLD	304
Dd		248	LSLAMDVIILGILLSGGTWASKHYK--NTNKSNSVACHFEPAIMAIIEHTPPLBEEKGRMEEMIN	305
OY		305	ETENLPRISEERVLVPPGDPERMANEYSOKAGGLVYOQCITALTEL	351
Dd		306	EIKNSKKHPFERIWHSEKGFLYMETKLKGIPIYK----KVDEL	348

RESULT 3	ID	MDH_MEMBER	STANDARD;	PRT;	339 AA.
AC	MDH_MEMBER	STANDARD;	PRT;	339 AA.	
AD	P16142:				
DT	01-APR-1990 (Rel. 14, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Malate/L-sulfolactate dehydrogenase (EC 1.1.1.37) (EC 1.1.1.82)				
DE	(Malate 1.1.1.272) ((R)-2-hydroxyacid dehydrogenase).				
GN	MDH.				
OS	Methanothermobacter fervidus.				
OC	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriales;				
OC	Methanothermobacteriaceae; Methanothermobacter.				
OX	NCBI_TaxID=2180;				
RP	[1]				
RP	SEQUENCE FROM N.A. AND SEQUENCE OF 1-24.				
RC	STRAIN=DSM 2088 / V24S;				
RX	MEDLINE=90235834; PubMed=2110059;				
RA	Honka E., Fabry S., Niemann T., Palm P., Hensel R.;				
RA	"Properties and primary structure of the L-malate dehydrogenase from				
RT	the extremely thermophilic archaeobacterium Methanothermobacter				
RL	Eur. J. Biochem. 188:623-632(1990).				
RL	[2]				
RP	FUNCTION.				
RP	MEDLINE=20309698; PubMed=10850983;				
RA	Graupner M., Xu H., White R.H.;				
RT	"Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing				
RT	reactions involved in coenzyme biosynthesis in methanocarchaea."				
RL	J. Bacteriol. 182:3688-3692(2000).				
CC	-1- FUNCTION: Acts on oxaloacetate, sulfolpyruvate but not on pyruvate.				
CC	Has a higher selectivity for the coenzyme NADH than for NADPH.				
CC	-1- CATALYTIC ACTIVITY: (S)-malate + NAD(P)(+) = oxaloacetate +				
CC	NAD(P)H.				
CC	-1- CATALYTIC ACTIVITY: (S)-3-sulfolactate + NAD(P)(+) = 3-				
CC	sulfolpyruvate + NAD(P)H.				
CC	-1- SUBUNIT: Homodimer.				
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.				
CC	-1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.				
CC	-----				
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X51714; CAA36010.1; -				
DR	EMBL; X51840; CAA36133.1; -				
DR	PIR; S08981; S08981.				
DR	InterPro: IPR005767; Idh_2.				
DR	Pfam; PF02615; Idh_2; 1.				
KW	Oxidoreductase; Tricarboxylic acid cycle; NAD; NADP.				
SQ	SEQUENCE 339 AA; 36762 MW; 23190822DB2/5835 CRC64;				

Query Match	24.0%	Score 465.5	DB 1	Length 339
Best Local Similarity	33.0%	Pred. No. 4.1e-31		
Matches 116	Conservative	67	Mismatches 150	Indels 15
			Gaps	8
QY	11	VSPKDAREFVVKQQTGVTSPPDHAGOLADLLDDADIVGHSHGLNRLHITVYDDVKNQNG-V	69	
DB	2	IISPEERSSLIITKILNALGVSEEHAKITADIVADLDKGFTHSGIGRFQYVYEGIKGTGI	61	
QY	70	KGNQVPAKVKOKGCTAWVDGENLGAIVGNFCTDIAIKIAKEFGVAWVVTKNSNHHYCAQO	129	
DB	62	KTSGNIEIEKETDSVALLNGHLLLOQVAAVKKMKALIAIEKAKNTGVGIVGIDHNSNHRGJAG	121	
QY	130	HYTKKIAAGAWGMSFTNTSPLMFRCRSSEIGLGNPLSCCNSKETSGLFLDMATTYV	189	
DB	122	YTSIDAMKNDMIGITMTETEPVAPVAGGKIPVGLNPIALISIPSNQY--YVAVDNSTAV	179	
QY	190	ALGKVELADCRGKTIPTSGWADSKGNPSTDTQVVLHGGLPLGIEETGSKYGTGLSM	249	
DB	180	ARGKL-LBAARKNEKIEPIGIAVDKKNGNPTTDPNEALN-GSLIPFG-----HKGYALCF	231	
QY	250	MGELFCGILAGSSPFGKSNVRLMGQSHKAADNGCAFVAIDQECF--APGFAPRLQOFLDFTNR	307	
DB	232	MIETLAGLVAAEFSGSKVGFVDPQOMCTKGDLILAIDPSKRYDIEEFKRVNDFEYKELK	291	
QY	308	NLKPISEKRPVLVPGDPERPMNTETYSQKAGGLVYQEGQIKALEFELATKCDYQM	359	
DB	292	-----STKDVAILPDGDRERMMIKKEKE-GELDKKLVLEIDDELINIEL	337	

```
CC -I- PATHWAY: Coenzyme M biosynthesis; third step.
CC -I- PATHWAY: Coenzyme methanoperlin diosynthesis.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE LPH2/MDH2 OXIDOREDUCTASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U67583; AAB99436..1; -.
DR PIR: H64477; H64477.
DR TIGR: MJ1425; -.
DR InterPro: IPR003767; Idh_2.
DR Pfam: PF02615; Idh_2; 1.
KW Coenzyme M biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 344 AA; 37358 MW; DDB9B086523CD298 CRC64;

Query Match          23.5%; Score 456.5; DB 1; Length 344;
Best Local Similarity 32.9%; Pred. No. 2,3e-30;
Matches 117; Conservative 64; Mismatches 156; Indels 19; Gaps 10;
```

YVSPDAREFYVKCMQTGVTSBDHAQDLADLLADVGHYSHGILRLHIYVDVKKNGVK

CC	COMM. MENUA	STANDARD;	PRT;	344 AA.
CC	COMC.METUA			
ID	Q58620;			
AC	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	L-sulfolactate dehydrogenase (EC 1.1.1.272) ((R)-sulfolactate			
DE	dehydrogenase)			
GN	COMC OR MDH OR M71425.			
OS	Methanococcus jannaschii.			
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;			
OC	Methanocaldococcaceae; Methanocaldococcus.			
OX	NCBI_TaxID:2190;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN-JAL-1 / DSM 2661 / ATCC 43067;			
RC	MEDLINE=96337999; PubMed=8688087;			
RX				
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,			
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,			
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,			
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,			
RA	Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,			
RA	Cottonback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,			
RA	Usterlund M.T., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,			
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;			
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus			
RT	jannaschii."			
RL	Science 273:1058-1073(1996).			
RN	[2]			
RP	FUNCTION.			
RP	MEDLINE=20309698; PubMed=10650983;			
RX				
RA	Graupner M., Xu H., White R.H.;			
RT	"Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing			
RT	reactions involved in coenzymal biosynthesis in methanocorinae."			
RL	J. Bacteriol. 182:3688-3692(2000).			
CC	-1- FUNCTION: Catalyzes the reduction of sulfolpyruvate to (R)-			
CC	sulfolactate much more efficiently than the reverse reaction. Also			
CC	catalyzes the reduction of oxaloacetate, alpha-ketoglutarate, and			
CC	to a much lower extent, KMGa, but not pyruvate. Involved in the			
CC	biosynthesis of both coenzyme M (with (R)-sulfolactate) and			
CC	methanopterin (with alpha-ketoglutarate).			
CC	-1- CATALYTIC ACTIVITY: (S)-2-sulfolactate + NAD(P)(+) = 3-			
CC	sulfolpyruvate + NAD(P)H.			

QY	71	GNVGV--VYLKOKGSTAVNDENILGAVNGPCDILAKIKENGAVNVMYKNSHVAC	128
Db	61	INPKDKIVATESPATVAIDBDLGLGVYVKKAMELAKKANKNGVGVATRNANHFSEA	120
QY	129	QHYTKLIAMAGVNGSMFTNSPLMFPCHSSEIGLGTNPLSCCVNSSEKGDSEFLDMATT	188
Db	121	GYSESLANQMDIGITITITTEPPAMAFGCKEIKLTNTPIALAFKGNK--KESIDMATAS	178
QY	189	VALGVELADORGKTOIPSTWGADSKSGNPSNDTQVYLHGGILPLGIGIEETSGYKGLS	248
Db	179	IARGKI-LEALRKTKRIPEGCAVDKDGKPTTIDPAKALE-GCILPPGG-----PKGYGLA	230
QY	249	MMGEILFCGILAGSSFEKGNVRLMGOSHKAAADQCFVAIDQECF--APGFAPRLQOFLDET	306
Db	231	LAIEMLSAI-NGAEGTKVKMGKTANPEEECTKGDFLAIINPEFMGKEEFRKRVBELDEI	289
QY	307	RNLKRISEKRVLYVNGDPERNTEISQKAGGLVYQEGIRKALELATCDVQMSY	362
Db	290	KNSEP-AGEFELLIPLETEERNK--MKRKDGEIDKNLYNQLKEICNELGINTEDY	342
RESULT 5			
COMC_METH			
ID	COMC_METH	STANDARD:	PRT: 341 AA.
AC	027273:		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	L-sulfolactate dehydrogenase (EC 1.1.1.-) ((R)-sulfolactate dehydrogenase).		
GN	COMC OR MTH1205.		
OS	Methanobacterium thermoautotrophicum.		
OC	Archaea: Euryarchaeota; Methanobacteria; Methanobacteriales;		
OC	Methanobacteriaceae; Methanothermobacter.		
OX	NCBI_TaxID=187420;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Delta H;		
RX	MEDLINE=98037514; PubMed=9371463;		
RA	Smith D.R., Doucette-Stamm L.A., Delonguey C., Lee H.-M., Dubois J.,		
RA	Aldehyde T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,		
RA	Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,		
RA	Spadafora R., Vitcare R., Wang Y., Wierzbowski J., Gibson R.,		
RA	Jiwani N., Caruso A., Bush D., Safer H., Patwell J., Pribnagar S.,		

RA McConaull S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Neelling J., Reeve J.N.,
RT "Complete genome sequence of *Methanobacterium thermoautotrophicum*
RT delta: functional analysis and comparative genomics." ;
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: Catalyzes the reduction of sulfolysinate to (R)-
CC sulfolactate. Involved in the biosynthesis of both coenzyme M
CC (with (R)-sulfolactate) and methanopterin (with alpha-
CC ketoglutarate) (By similarity).
CC -1- CATALYTIC ACTIVITY: (R)-sulfolactate + NAD(+) = sulfolysinate +
CC NADH.
CC -1- PATHWAY: Coenzyme M biosynthesis; third step.
CC -1- PATHWAY: Coenzyme methanopterin biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE LHD2/MDH2 OXIDOREDUCTASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on lists
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib.ch).
CC
CC -----
DR EMBL, AE000888, AAB85694.1; -
DR PIR, H69027; H69027.
DR InterPro: IPR003767; Idh_2.
DR Pfam: PF02615; Idh_2; 1.
KW Coenzyme M biosynthesis; Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 341 AA; 36630 MW; AC266E12401A1184 CRC64;

Query Match	21.0%;	Score 407;	DB 1;	Length 341;
Best Local Similarity	29.6%;	Pred. NO. 2.8e-26;		
Matches 104;	Conservative 77;	Mismatches 150;	Indels 20;	Gaps 8;

QY	12	USPDAAREVYKCMQOTVSTDPHQAQLDLILDADLYGSHGLNLTHTYVDYKNG--	70
Dd	3	ISPEEVKVIIEILAMNPRESSIVADYIDADKGFSSHGIGFPQIVDILRRGTT	62
QY	71	GNVGPVKYLQKGGTAMVDGENLLGAVNGFTDLAIKLBKRGVAVWVTKNSNHGACOH	130
Dd	63	ADGGITITERESTPLINGNHIFGVAVYRAEMLIEKARNNGVGLGVGHDSNHFVAGY	122
QY	131	YTKRIANGVMQSTNTSPLMPCORSSEIGGTAPLDCSCVSEKSGDGFILDMATTYA	190
Dd	123	KSDAMVMDMIGVIANTEPAVAPVLCGRKPIILGTNPVALGIFSNKRYVS--	180
QY	191	LKGYELADCRGKTQIIPSTWGAADSKGNPSTDYVVLHGGLPLGCIETGYSYKGLSKM	250
Dd	181	RKGLLEARKKES--IPENVALDAEKKPTTDPEMALEK--GSILTFPG--	232
QY	251	GELFCGIIAGSGFNGVNRIMWGSHRAADNGCFCFVAIDCECPA--PGFAPRLOOFLETRN	308
Dd	233	IEILAGPLVGAAGFAVYGTANPEEMCKRGDMLAIDPKMNDPDEFRAQVDFEIE---	289
QY	309	LKPISEKRPVLPGRPERKNTFYISQKAGGLVYQEOIKALBELATACDYOM	359
Dd	290	---VKSSGVLDIPGIESMNIK--RRRAQIEIDEXLEKLEETILIAARLDINL	336

RESULT 6	
YJMC_BACSU	
ID	YJMC_BACSU
AC	034736;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Hypothetical oxidoreductase yjmc (EC 1.1.1.-).
GN	yjmc.
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus
OX	NCBI_TaxID=1423;
RN	[1]

RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA MEDLINE-98240235; PubMed-9579062;
RX Rivolta C., Soldo B., Iazarevic V., Joris B., Manuel C., Karamata D.;
RT "A 35.7 kb DNA fragment from the *Bacillus subtilis* chromosome
RT containing a putative 12.3 kb operon involved in hexonate catabolism
RT and a perfectly symmetrical hypothetical catabolite-responsive
RT element.";
RL Microbiology 144:877-884(1998).

RP SEQUENCE FROM N.A.

RX MEDLINE=98044033; PubMed=9384377

RA Azevedo V., Bertero M.G., Bessleres P., Bolotin A., Borchert S.,

RA Brouillet S., Bruschì C.V., Caldwell B., Capuano V., Carter N.M.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,

RA Fritzt C., Fujita M., Fuma S., Galizzi A., Galleron N.,

RA Gulsepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kurita K., Lapidus A., Lardino S., Lauber J., Lazarevic V.,

RA Medina N., Mellado B. P., Mizuno M., Moestl D., Nakai S., Noback M.,
Dec D. M., Devline A., Liu H., Masuda C., Mucci C., Neuzil C.,

BA Parro V. Pohl T M. Portetelle D. Portwollik S. Prescott A M.
NA Nounie L., Uygawa K., Uygawa A., Uygawa B., Uygawa S.H.,
NA Nounie L., Uygawa K., Uygawa A., Uygawa B., Uygawa S.H.,

KA Priesecan E., Pujic F., Fumelle B., Rapoport G., Key M., Keynorius S
BA Plescer M, Pivolta C, Roche B, Rose M, Sadate Y

RA SATO T.; SCANDIAN E.; SCHROETER K.; SCOLLIONE F.,
PA SOKIYACHI T. SOKRUSKA A. SORRIS I. SHIN B S. SOLDI

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takahashi Y., Tashiro M., Tawada D., Tawara T., Tawara Y., Tawara Z.,
Tawara A., Tawara B., Tawara C., Tawara D., Tawara E., Tawara F., Tawara G.,
Tawara H., Tawara I., Tawara J., Tawara K., Tawara L., Tawara M., Tawara N.,
Tawara O., Tawara P., Tawara Q., Tawara R., Tawara S., Tawara T., Tawara U.,
Tawara V., Tawara W., Tawara X., Tawara Y., Tawara Z.

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*

RL Nature 390:249-256(1997).

CC -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collabora

CC the European Bioinformatics Institute. There are no restrictions on

CC modified and this statement is not removed. Usage by and for comm

CC or send an email to license@isb-sib.ch).

DR EMBL: AF015825; AAC46328.1; -.

DB PTB: E69852: E69852.

DA subc1e1s1c, b6132v0, yjmc.
 DR Interpro: TPB003767: 1db 2

DK P1aIII; PF02613; IGL2; 1.
 KW Hypothetical protein: Oxidoreductase: Complete proteome

SEQUENCE 337 AA; 364/0 MW; C282CUA8F1L3FEAC CRC64;

Query Match	Score	DB	Length
20.18;	390.5;	DB 1;	33/;

Matches 104; Conservative 68; Mismatches 142; Indels 15; Gaps

12 VSPKDAREFVKCMQTGTS PDHAGQLADLLDADLVGHYSHGLNRLHIYVDDVKNGVH

Db 6 IAAEEAKELVQKLDGAGLNERDAEKVADVLVHADLRNVHSHGVLHTEHYVNRLLAGG

72 NGVPKVLKQGG-TAWVDGENLTGAVVGNFCTDLAIKLAKEFGVAVVTKNSNHYGACQ

```

Db      66  PGADPVEKETGPTVGLDDGDFGVHNCDDMADHIDMAKKKGVMTAVNVSHCALSY 125
Qy      131 YTKRIANAGMVGMSFTNTSPLMPCRSSEIGLTNPISCCVNSEGTGSLDMATTVA 190
Db      126 FVQAAADKLGLGMATHTDSIVVPGGRTPLGTNPDIAYGVA--KHKKPFLDMATSRVA 184
Qy      191 LGKVELADCRGKTQIPSTWGAADSKGNPSTDTQVYVHGGLPLGIGIEETGSYKGTGLSM 250
Db      185 FGRILQAREBEK-ELPEGMGVDENGEAVTDPKV-----SLSTFGGPKGGLSTIV 234
Qy      251 GELFCGILAGSSFGKNV-RLWGQSHKADNGCCFVAIDQCFPA--PGFAPRLQDFLDETR 307
Db      235 VDVEGSLAGAFGPHIAKMYNGLDQKRKLGHVYCAINPSPFTDFTLEQDMADIDELQ 294
Qy      308 NLKPTSEKPYLVPGDEPERMNTYSQKAG 336
Db      295 QSPPAVGFERVYVPEIEIOLHEERKKNG 323

```

RESULT 7

```

ALD_EC057
ID      ALD_EC057      STANDARD:      PRT:      349 AA.
AC      P58408;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DE      Ureidoglycolate dehydrogenase (EC 1.1.1.154).
GN      ALD OR GLX8 OR Z0672 OR EC50579.
OS      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Escherichia.
RN      NCBI_TaxID=83334;

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX      MEDLINE=21074935; PubMed=11206551;
RA      Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA      Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA      Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamoustis K.,
RA      Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA      Welch R.A., Blattner F.R.;
RT      "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL      Nature 409:529-533(2001).

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / RMD 0509952;
RX      MEDLINE=21156231; PubMed=11258796;
RA      Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA      Ikida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA      Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT      "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12.";
RL      DNA Res. 8:11-22(2001).
CC      -1- FUNCTION: INVOLVED IN THE ANAEROBIC UTILIZATION OF ALLANTOIN.
CC      -1- CATALYTIC ACTIVITY: (S)-ureidoglycolate + NAD(P)(+) = oxaluate +
CC      NAD(P)H.
CC      -1- PATHWAY: Degradation of allantoin (purine catabolism) in the
CC      direction of oxamate formation (third step.
CC      -1- SUBCELLULAR LOCATION: CYTOPLASM (Probable).
CC      -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb-sib.ch).
DR      EMBL; AE005232; AAC54874.1; -

```

```

DR      EMBL; AP002552; BAB34002.1; -
DR      PIR; C90701; C90701.
DR      PIR; F85551; F85551.
DR      InterPro; IPR003767; Idb_2.
KW      Pfam; PF02615; Idb_2; 1.
KW      Oxidoreductase; NAD; Purine metabolism; Complete proteome.
SQ      SEQUENCE 349 AA; 37891 MW; 2759E09DBD10C964 CRC64;

```

Query Match 18.6%; Score 360.5; DB 1; Length 349;
 Best Local Similarity 31.8%; Pred. No. 26-22;
 Matches 102; Conservative 49; Mismatches 149; Indels 21; Gaps 10;

```

Qy      12 VSPKDAREFVYVCMQTVGSPDHAGOLADLLDADLVGYSHGLNRHLHYVDVNGKVG 71
Db      3 ISRETLHOLIENKLIQAGIKREHATVAEVLVYADARGIHSGAVRVEYAIRISK-GT 61
Qy      72 NGVPVFLKQKG--TAVVDGENILGAVVNGFCTDLAIKLEFGVAVVYTKNSNYGACQ 129
Db      62 NREPEFLTEETGPCSAILHADNAGOVAAKMGMEHAIKTAQNGVAVVGISMGSGAIS 121
Qy      130 HTKRIANAGMVGMSFTNTSPLMPCRSSEIGLTNPISCCVNSEKTGDSFL-LDMATT 188
Db      122 YFVQAAARAGLIGISMCQSDPMVVPFGGEIYGTNPILAFAPGE--GDEILTFDMATT 179
Qy      169 VALGVELADCRGKT-QIPSTWGAADSKGNPSTDTQVYVHGGLPLGIGIEETGSYKGTGL 247
Db      160 QAMGV--LDARSRMKSTIPDTAVYKNCAPTTPRAV---HALLEPAAG-----PKYIGL 228
Qy      248 SMWGLFCGILAGSSFGKNV-RLWGQSHKADNGCCFVAIDQCFAPG--FAPRLQDFLD 304
Db      229 MMIMIVLSGVLLGLPFGRQVSSMYDDLHAGRMIGQLHVYINPFPSSSELFRHLSQTM 288
Qy      305 ETRNLKPTSEKPYLVPGDPE 325
Db      289 ELNATTPAGFENQVYVPGQDQ 309

```

RESULT 8

```

ALD_EC001
ID      ALD_EC001      STANDARD:      PRT:      349 AA.
AC      P77555;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DE      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Ureidoglycolate dehydrogenase (EC 1.1.1.154).
GN      ALD OR GLX8 OR B0517.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Escherichia.
RN      NCBI_TaxID=562;

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / EC11;
RX      MEDLINE=20069628; PubMed=10601204;
RA      Cusa E., Odradors N., Baldoma L., Badia J., Aguilari J.;
RT      "Genetic analysis of a chromosomal region containing genes regulated
RT      for assimilation of allantoin nitrogen and linked glyoxylate

```

```

RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / EC11;
RX      MEDLINE=20069628; PubMed=10601204;
RA      Cusa E., Odradors N., Baldoma L., Badia J., Aguilari J.;
RT      "Genetic analysis of a chromosomal region containing genes regulated
RT      for assimilation of allantoin nitrogen and linked glyoxylate

```


DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical oxidoreductase yb1c (EC 1.1.1.-).
 GN yb1c OR B0801.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=94312114; PubMed=8037924;
 RA Ohmori H.;
 RT "Structural analysis of the rhlE gene of Escherichia coli.";
 RL Jpn. J. Genet. 69:1-12(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=9905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamuro Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: L02123; AAA53657.1;
 DR EMBL: AE000182; AAC73888.1;
 DR EMBL: D90717; BA35467.1;
 DR PIR: A64817; A64817.
 DR Ecogene: EG11581; yb1c.
 DR InterPro: IPR003767; ldh_2.
 DR Pfam: PF02615; ldh_2; 1.
 DR Hypothetical protein: Oxidoreductase; Complete proteome.
 FT CONFLICT 23 24 MG -> IP (IN REF. 1).
 SQ SEQUENCE 361 AA; 38897 MW; 2829784979E48543 CRC64;
 Query Match 16.8%; Score 326.5; DB 1; Length 361;
 Best Local Similarity 29.2%; Pred. No. 1.3e-19;
 Matches 100; Conservative 59; Mismatches 161; Indels 23; Gaps 11;
 OY 20 FVVKCMQVGTSPHAGQADLLDADLVGHYSHGLNRLHYVDVKNK-VKNGVPRVL 78
 DB 15 FIAVFRQMGSEDEAKLVADHLIAANLAGHDSHGICIPSVRSWMSGHIOINHAKT 74
 OY 79 KQKGTAMVDENLGAIVNGFCIDLAIKLAKGVAVVYTKNSNHGACOHYTKKIANA 138
 DB 75 KEAGAAVYLDGDRAFGQVAHEAMALGIEKAHGHIAVAVALNSHHIRIGYVWQCAA 134

OY 139 GMVGSFNT--SPLMPFCSRSEIGLTNPISCYVNSKRGDPSFLDMATTTVALKVEL 196
 DB 135 GFVSHFVSVYGIPIVAPFHHGDSRFGRNPF-CVYFPRKNDPPLLDVATSAIAGKTRV 193
 OY 197 ADCRGKTIPISTWADSKGNSTDTQVYLHGGGLPLCGIEFTGSKYKGTGLSMGELFCG 256
 DB 194 AWHKG-VVPVPGCLIDVNGVPTNPVWQES---PLGSLTLFAHKKVYALAAWCEILGG 248
 OY 257 ILAGSFGKRV--RLMGSHKADAGQCFVADIDECF-APGAPRRLQGFDETRLRKIS 313
 DB 249 ALSG--GKTHQETLQTSPPAILNCMTTIIINPELPAPDCNMQTEFAFVWK-ASPHD 304
 OY 314 EEKPLVPGDPERMNTFYSOKAGGLVYEGQIKALEEATKCD 356
 DB 305 DDKPLPLGEMF-VNTRRERKOGIPLDAGSQAL-----CD 340
 RESULT 11
 YIAK_ECOLI
 ID YIAK_ECOLI STANDARD; PRT; 332 AA.
 AC P37672; P76716;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical oxidoreductase yiaK (EC 1.1.1.-).
 GN yiaK OR B3575.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 region from 76.0 to 81.5 minutes.";
 RL Nucleic Acids Res. 22:2576-2586(1994).
 RN [2]
 RP SEQUENCE OF 1-61 FROM N.A.
 RA Badia J., Ibanez E., Sabate M., Baldoma L., Aguilar J.;
 RT Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U00039; AAB18552.1;
 DR EMBL: AE000435; AAC76599.1;
 DR EMBL: AJ223475; CAA11398.1;
 DR PIR: A65157; A65157.
 DR Ecogene: EG12279; yiaK.
 DR InterPro: IPR003767; ldh_2.
 DR Pfam: PF02615; ldh_2; 1.
 DR Hypothetical protein: Oxidoreductase; Complete proteome.
 SQ SEQUENCE 332 AA; 36572 MW; A411EB14D5C03DF7 CRC64;
 Query Match 13.9%; Score 269.5; DB 1; Length 332;
 Best Local Similarity 26.1%; Pred. No. 6.2e-15;
 Matches 79; Conservative 64; Mismatches 133; Indels 27; Gaps 8;
 OY 29 GTSPPHAGQADLLDADLVGHYSHGLNRLHYVDVKNK-VKNGVPRVLKQKGTAMV 87
 DB 20 GVDSETADCAEMFARTTESGYSHGVNFPFIOOLENGDIIIPADPKRTISLCAIOW 79
 OY 88 DGENLLGAVVNGFCIDLAIKLAKGVAVVYTKNSNHGACOHYTKKIANGMVGSFNT 147

```

Db      80 DAQRSGNLTAKKMDRALELAADHGIGVALRNNHMRGSGYGMQAEKGYIGLCWTN 139
QY      148 TSDLPFPCRSSEIGLGNPLSCCNSEKTSDFLLDMATTVALGKVELADCKGKTIIPS 207
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      140 SIAMFPMWAKKEKRIGNPLIYAIIPSPIT---TWDMKSMSPFYGMLEVRNLAGR-OLPV 195
QY      208 TWGADSKNDSTDTQVVLHGGGLPLUGLEETGYSYKGTGLSMGELFCGILAGSSFEKGV 267
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      196 DGGFDEGNLTKEPVEYKRNRIIPM-----GYWKGSGMSIYLDMIATL---SDGASV 246
QY      268 RLWGSGHKAADN-----GGCFVAI---DOECFAPGFAPRLQOFLDERNPKPISEKPV 319
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      247 -----AEVTDNDEIGISQIFAIYDKLIDGPTRDAKLQRIIMDYTSARADENQAIR 301
QY      320 VPG 322
        : : :
Db      302 LFG 304

RESULT 12
LDH_ALCEU
ID      LDH_ALCEU      STANDARD;      PRT;      349 AA.
AC      007251;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      L-lactate dehydrogenase (EC 1.1.1.27).
GN      LDH.
OS      Alkaligenes eutrophus (Ralstonia eutropha).
OC      Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
        Burkholderiaceae; Ralstonia.
OX      NCBI_TaxID=510;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.
RX      STRAIN-H16 / DSM 428 / ATCC 17699;
        MEDLINE=94010251; PubMed=8405966;
RA      Jendrossek D., Kratzin H.D., Steinbuechel A.;
RT      "The Alkaligenes eutrophus ldh structural gene encodes a novel type
        of lactate dehydrogenase."
RL      FEMS Microbiol. Lett. 112:229-235(1993).
CC      -1- CATALYTIC ACTIVITY: (S)-lactate + NAD(+) -> pyruvate + NADH.
CC      -1- PATHWAY: Anaerobic glycolysis; final step.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC      -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
        or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: Z22737; CAAB0432.1; -
        PIR: I39535; I39535.
DR      InterPro: IPR003767; ldh_2.
DR      Pfam: PF02615; ldh_2; 1.
KW      Oxidoreductase; NAD; Glycolysis.
SQ      SEQUENCE 349 AA; 36755 MW; 7A8FB0E7859CBC04 CRC64;

Query Match      13.8%; Score 268; DB 1; Length 349;
Best Local Similarity 28.2%; Pred. No. 8.8e-15;
Matches 103; Conservative 53; Mismatches 167; Indels 42; Gaps 13;

QY      12 VSPKDAREFVVKCMQTVGTSPDHAGLADLLDADLVGHSHGLNRLHIYVDVK-NGVK 70
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      3 ISLTSARQLARLDLAQVPADLADVDVAHELVESDCKGYISHLSTLIPNRTKLDGHSVN 62
QY      71 GNGVPEVLKQGGTANVDENLIGAVVNEFCYDLAIKLAKFEGVAVVYTKNSNHYGACOH 130
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      63 PGSRACVLDGGLTAFEDDGGFGQVHGKSVQAALIERVRQHGHCIVTLRRSHHLGRMGH 122
QY      131 YRKIKNANMGVMSFTNT---SPLMFPCRSSEIGLGNPLSCCNSEKIG-DSFLDMAT 186

```

```

Db      123 YGEMAAAGCVLLSFTVIVNRAVPVAFGGVRAVRLFTTNPL-CEAGPMNPRLPVYDIAT 181
QY      187 TYVALGVELADCKR-TOIPSTWGSADSKNPSTDPQVVL--HGGLLPLGIEERGSVK 243
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      182 SAIAINARVARLAKRGEAPGCSITIGAD--GNPTDASTMGHERPGALPPGG-----HK 233
QY      244 GTGLSNMGELFCGILAGSSFGKVVRLMGOSHKRAADNGOCFVAIDOCFAPGFAPRL----- 299
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      234 GVALGVYAEELLAGVLSG---GGTIQ-----PDNPRGVAVN-MLFVLLNPAIDLGL 281
QY      300 -----QOFLDTRNPKPISEKPVLPYGPDPERRMNTYSQKAGGLVYQSGIKALELA 352
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      282 DMQSAEYEAFFVRLHPTPAPGVDRVQYGPGEYEAANR-AQASDTININPAIWRNRLA 339
QY      353 TKCDV 357
        : : :
Db      340 QSLNV 344

```

```

RESULT 13
YC56_PASMU
ID      YC56_PASMU      STANDARD;      PRT;      332 AA.
AC      Q9CLH5;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Hypothetical oxidoreductase PM1256 (EC 1.1.1.-).
GN      PM1256.
OS      Pasteurella multocida.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
        Pasteurellaceae; Pasteurella.
OX      NCBI_TaxID=747;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-Pm70.
        MEDLINE=21145866; PubMed=11248100;
RA      May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT      "Complete genomic sequence of Pasteurella multocida Pm70."
RL      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC      -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
        or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AE006164; AAK03340.1; -
        InterPro: IPR003767; ldh_2.
DR      Pfam: PF02615; ldh_2; 1.
KW      Hypothetical protein; Oxidoreductase; Complete proteome.
SQ      SEQUENCE 332 AA; 36868 MW; CE6D4A45C87CF905 CRC64;

Query Match      13.3%; Score 258; DB 1; Length 332;
Best Local Similarity 26.4%; Pred. No. 5.5e-14;
Matches 96; Conservative 65; Mismatches 153; Indels 50; Gaps 14;

QY      1 MNTSKDAPEP--VSPKDAREFVVKCMQTVGTSPDHAGLADLLDADLVGHSHGLNRL 58
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      3 VSYEALKQEFKRVLLARNVRE-----DIAECATMFADTTESGYVSHGVNRF 49
QY      59 HIYVDVYKNGVNGVPE-----KVLKQGGTANVDENLIGAVVNEFCYDLAIKLAKERG 113
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      50 PRETSOLE---KGDVPADEPTVLSLGAIEQR-DHQAIQNTAKKMPDRAEELIDQFG 105
QY      114 VAVVYTKNSNHYGACOHYTKIKIANMGVMSFTNTSPLMFPCRSSEIGLGNPLSCCVNS 173
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      106 IGVVALKNNHNMRRGGYGMQAAEKGYIGLCWTNSIAVMPWAKKRCRGTNPLLIAY-- 163

```


QY 174 EKTGDSFLDMATTVALGKVELADCRGKTQIPSTWAGDSKNSPTDQVYLHGGLPL 233
 DB 164 -PTPTPTWDMSCMSYSGMELVHRLAQRQTFVDA-GRDDDEGNLDRDGTVEKNRLLPM 221
 QY 234 GGIEETGSKGTGSLMMGELFCGLIAGSSFGKNVRLMGQSHKADNGOC---EVAIDOE 289
 DB 222 -----GFWKSGSLSIYLDMTATLL---SNGLSV---AEVTEEKDEYCVSQIFTAIEVD 269
 QY 290 CFAPGFA--PRLOQFLDDETRNLKPISEKPYLVGDPDEPRMTEYSQKAGLVYQE---GQ 344
 DB 270 RLIDGNVTDEKLNKIMDYRTAERADPDVAIRLPQH-EFTAIRKKNKANGIPDETWEK 328
 QY 345 IKAL 348
 DB 329 IKSL 332

RESULT 14
 YIAK_HAEIN STANDARD: PRT; 332 AA.

AC P44995:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical oxidoreductase HI1031 (EC 1.1.1.-).
 GN HI1031.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McEweney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utermbeck T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
 RA Fine L.D., McDonald L.A., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
 CC -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U32783; AAC22691.1; -.
 DR PIR: C64165; C64165.
 DR TIGR: HI1031; -.
 DR InterPro: IPR003767; Idb_2.
 DR Pfam: PF02615; Idb_2; 1.
 KW Hypothetical protein; Oxidoreductase; Complete proteome.
 SO SEQUENCE 332 AA; 36970 MW; FCFABG66ZFE495FBF CRC64;

Query Match 13.3%; Score 257.5; DB 1; Length 332;
 Best Local Similarity 25.4%; Pred. No. 6,1e-14;
 Matches 76; Conservative 56; Mismatches 136; Indels 31; Gaps 7;

QY 35 AGQADLLDADLVGHSHGNRLHITYDDVYKNG-VKNGVYPKYLKQGGGTAWDGENLL 93
 DB 26 AECATATPTDTTQAGAVSHGINRPRTIQOLEOGDIVNAIPTKVLISGSIEMOWAHQAI 85

QY 94 GAVVGNFCTDLAIKLAKEFGVAVVYTKNSNHYGACQHTKRIANAGVMSFTNTPLMF 153
 DB 86 GNLTAKKMMADRIELASGHGVIALRANHMNRGSGWGMAAEKGIYGIOMTNLAAMP 145
 QY 154 PGRSEIQLGNIPLSCVNSEKTDGSLDMATTVALGKVELADCRGKTQIPSTWAGDS 213
 DB 146 PWGAKERCIRGTNPILIAV---PTPTPTWDMSCMSYSGMELVHRLAQRQTFVDA-GRDD 201
 QY 214 KGNPSTDQVYLHGGLPLGIEFTGSKGTGSLMMGELFCGLIAGSSFGKNVRLMGQS 273
 DB 202 EGNLTRDPSIVEKNRLLPM-----GFWKSGSLSIYLDMTATLLSN-----GES 245
 QY 274 HKAADN-----GQCFVAIDQECFAGFA--PRLOQFLDDETRNLKPISEKPYLVG 322
 DB 246 TVAVTEDEKNDKCYQVFAIEVDRLIDGKSKDEKLNIMDYKTAERSQGIAMAVNRCLPG 304

RESULT 15
 YC52_PSEAE STANDARD: PRT; 334 AA.

AC Q91492:
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical oxidoreductase PA1252 (EC 1.1.1.-).
 GN PA1252.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
 CC -1- SIMILARITY: BELONGS TO THE LDH2/MDH2 OXIDOREDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE004555; AAG04641.1; -.
 DR PIR: B83488; B83488.
 DR InterPro: IPR003767; Idb_2.
 DR Pfam: PF02615; Idb_2; 1.
 KW Hypothetical protein; Oxidoreductase; Complete proteome.
 SO SEQUENCE 334 AA; 35683 MW; EFG9A6BA407515A7E CRC64;

Query Match 11.7%; Score 227.5; DB 1; Length 334;
 Best Local Similarity 27.0%; Pred. No. 1,8e-11;
 Matches 76; Conservative 48; Mismatches 131; Indels 27; Gaps 9;

QY 16 DAREFVVCQMOTVGTSPDHAGQLADLLDADLVGHSHGNRLHITYDDVYKNG-VKNGV 74
 DB 8 EYRELAIVLIRRHARSEAHVQAVADTLVAGNDEBACSHGIRLGCITLAKAGKSAAE 67
 QY 75 PKVLKQGGGTAWDGENLLGAVVGNFCTDLAIK-----AKEFGVAVVYTKNSNHYG 126
 DB 68 PELHDIAGLLRVVDAN-----GGF-SQCAFRLGLPHILRKARSQGIAMAVNRCLVFS 119

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:15:35 ; Search time 86 Seconds

(without alignments)
675.511 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 366
Sequence: 1 MNYSKDAPEFVSPKDAPEF.....ALEELATKCDVQMFYKRLK 366

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.5	220	22	AAO13791
2	9	2.5	226	22	ABBI7770
3	8	2.2	446	22	ABB65105
4	8	2.2	474	24	ABU11400
5	7	1.9	38	21	AAB51764
6	7	1.9	61	22	AAM84417
7	7	1.9	75	22	ABB03297
8	7	1.9	75	24	ABU12591
9	7	1.9	76	21	AAB54024

10	7	1.9	83	21	AAG02242	Human secreted pro
11	7	1.9	90	22	AAE10715	Human 4S13GalIIH do
12	7	1.9	90	22	AAE10716	Mouse 4S13GalIIH do
13	7	1.9	90	22	AAE10717	Pig 4S13GalIIH doma
14	7	1.9	91	22	ABG59631	Human liver peptid
15	7	1.9	91	22	ABB44259	Peptide #11765 enc
16	7	1.9	91	22	ABB27130	Protein #9129 enc
17	7	1.9	91	22	AAM65297	Human brain expres
18	7	1.9	91	22	AA21888	Peptide #8322 enco
19	7	1.9	91	22	AA38214	Peptide #12251 enc
20	7	1.9	104	23	ABP60785	Rhodospirillum rub
21	7	1.9	106	23	ABP08476	Human ORFX protein
22	7	1.9	125	22	AAU65451	Propionibacterium
23	7	1.9	137	22	AAU01885	M. tuberculosis an
24	7	1.9	146	19	AA46785	Yeast NC2-beta/DRI
25	7	1.9	146	22	AAW93521	Human polypeptide,
26	7	1.9	146	23	ABW06782	Human cancer cell
27	7	1.9	146	23	ABW04727	Human PPI143 prote
28	7	1.9	148	18	AAW89762	Staphylococcus aur
29	7	1.9	151	22	AAE01517	Human gene 2 encod
30	7	1.9	161	22	AAU09080	Human CDNA tyrosin
31	7	1.9	174	22	AAU09079	Human tyrosine pho
32	7	1.9	180	22	AAE01437	Human gene 2 encod
33	7	1.9	180	23	ABG63866	Human albumin fusi
34	7	1.9	195	22	AAE01465	Human gene 2 encod
35	7	1.9	195	23	ABG63863	Human albumin fusi
36	7	1.9	199	24	ABP79001	N. gonorrhoeae ami
37	7	1.9	209	22	AAE01516	Human gene 2 encod
38	7	1.9	210	20	AAV60556	Human normal blad
39	7	1.9	213	22	AAW90922	C glutamicum prote
40	7	1.9	221	23	ABW47829	Listeria monocytog
41	7	1.9	229	22	AAU34290	Staphylococcus aur
42	7	1.9	230	22	AAU37191	Staphylococcus aur
43	7	1.9	230	22	AAU37459	Staphylococcus epi
44	7	1.9	231	23	ABP39249	Staphylococcus epi
45	7	1.9	242	22	AAU01883	M. tuberculosis an

ALIGNMENTS

RESULT 1	AAO13791	standard; Protein; 220 AA.
ID	AAO13791	
AC	AAO13791:	
DT	06-NOV-2001	(first entry)
DE	Human polypeptide SEQ ID NO 27683.	
XX	Human: cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vacuole; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorders; arthritis; inflammation.	
OS	Homo sapiens.	
PN	WO200164835-A2.	
XX		
PD	07-SEP-2001.	
XX		
PF	26-FEB-2001; 2001WO-0504927.	
XX		
PR	28-FEB-2000; 2000US-0515126.	
XX		
PA	18-MAY-2000; 2000US-0577409.	
XX		
PI	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Drmanac RT;	
XX		
DR	WPI; 2001-514838/56.	
DR	N-PSDB; AA193722.	

XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PS Claim 20; SEQ ID NO 27683; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
CC cytokine cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 220 AA;
Query Match 2.5%; Score 9; DB 22; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 346 KALELATK 354
|||||||
DB 121 KALELATK 129
RESULT 2
ABBI7770
ID ABBI7770 standard; Protein; 226 AA.
XX
AC ABBI7770;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 6427.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virocidic;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
KW antiparasitic; hepatotropic; cerebroprotective; anti-inflammatory;
KW antiallergic; antidiabetic; antidiabetic; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
PD 16-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.

01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI N-PSDB; ABA14096.
 DR
 XX
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 XX
 PS Claim 11: SEQ ID NO 6427; 1701pp + Sequence Listing; English.
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral ischaemia;
 CC and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 226 AA;
 Query Match 2.5%; Score 9; DB 22; Length 226;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 346 KALELATK 354
 Db 127 KALELATK 135
 RESULT 3
 ID ABB65105 standard; Protein: 446 AA.
 XX ABB65105;
 AC ABB65105;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 22107.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-6556860/75.
 DR N-PSDB; ABL09208.
 XX
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 22107; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 446 AA;
 Query Match 2.2%; Score 8; DB 22; Length 446;
 Best Local Similarity 100.0%; Pred. No. 28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 180 FLDMMATT 187
 Db 262 FLDMMATT 269

RESULT 4
 ABU11400
 ID ABU11400 standard; Protein: 474 AA.

XX AC ABU11400;
 XX DT 11-FEB-2003 (first entry)
 XX DE

XX Protein encoded by S. atroolivaceus leinamycin gene cluster ORF lmmx.
 XX KW Leinamycin biosynthesis gene cluster; lmm; open reading frame; ORF;
 KW anti-tumour antibiotic; broad spectrum antimicrobial activity;
 KW Gram-positive; Gram-negative bacteria; chemical modification;
 KW metabolite; apo-carrier protein; holo-carrier protein; tumour;
 KW polypeptide; hybrid polypeptide/polypeptide metabolite; lmm production;
 KW cytosolatic.

XX OS Streptomyces atroolivaceus.
 XX PN WC200277179-A2.
 XX PD 03-OCT-2002.

XX PF 22-MAR-2002; 2002WO-US08937.
 XX PR 26-MAR-2001; 2001US-278935P.
 XX PS

PA (REGC) UNIV CALIFORNIA.
 PA (KYO) KYOMA HAKKO KOCYO KK.
 PI Shen B, Cheng Y, Tang G;

DR WPI: 2003-018907/01.
 DR N-PSDB: ABX34289.
 XX

PT Novel gene cluster responsible for synthesis of leinamycin in
 PT Streptomyces atroolivaceus useful for making various peptide and/or
 PT polypeptide, and/or hybrid polypeptide/polypeptide metabolites

XX PS Claim 13; Page 152-153; 185pp; English.

CC The present invention relates to the isolation of the Streptomyces
 CC atroolivaceus leinamycin (lmm) biosynthesis gene cluster containing
 CC 71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lmma through
 CC lmm2, and ORFs +1 through +9). Leinamycin is a novel anti-tumour
 CC antibiotic produced by several Streptomyces species. It exhibits
 CC broad spectrum antimicrobial activity against Gram-positive and
 CC Gram-negative bacteria, but not against fungi. The polypeptides encoded
 CC by the lmm biosynthesis gene cluster ORFs are useful for chemically
 CC modifying a molecule in a host cell. The host cell is a bacterium or
 CC eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect
 CC cell. The molecule is an endogenous metabolite produced by the host
 CC cell or exogenously supplied metabolite, or an amino acid, and the
 CC polypeptide is a peptide synthetase or amino transferase. The
 CC polypeptides encoded by the lmm gene cluster are useful for converting
 CC an apo-carrier protein to a holo-carrier protein. lmm shows potent
 CC antitumour activity in tumour models in vivo. The lmm gene cluster
 CC modules and/or catalytic domains are useful for making various peptide
 CC and/or polypeptide, and/or hybrid polypeptide/polypeptide metabolites.
 CC The proteins encoded by the ORFs are useful alone, or in combination
 CC with other active domains to modify various target substrates. The
 CC lmm gene cluster is useful to upregulate endogenous lmm production to
 CC permit lmm production in cells and/or to make various modified lmm.
 CC lmm, its analogue, or other polypeptide, peptide or hybrid
 CC polypeptide/polypeptide metabolites are useful as therapeutic agents, to
 CC treat a number of disorders, depending upon the type of metabolites.

CC ABU11341-ABU11411 represent the proteins encoded by ORFs of the
 CC S. atroolivaceus leinamycin biosynthesis gene cluster.
 XX

SO Sequence 474 AA;
 Query Match 2.2%; Score 8; DB 24; Length 474;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 GGLPLPG 235
 Db 74 GGLPLPG 81

RESULT 5
 AAB51764
 ID AAB51764 standard; Protein: 38 AA.

XX AC AAB51764;
 XX DT 16-FEB-2001 (first entry)
 XX DE

XX Human secreted protein sequence encoded by gene 35 SEQ ID NO:93.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; vulnerary; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging.

XX OS Homo sapiens.
 XX PN WC2000061625-A1.
 XX PD 19-OCT-2000.

XX PF 06-APR-2000; 2000WO-US08981.
 XX PR 09-APR-1999; 99US-0128701.
 XX PR 20-JAN-2000; 2000US-0177166.

XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PA (ROSE/) ROSEN C A.
 PI Rosen CA, Ruben SM, Komatsoulis G;

XX WPI: 2000-619226/59.
 XX DR N-PSDB: AAC93456.
 XX

PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives

XX PS Claim 11; Page 460; 500pp; English.

CC Polynucleotide sequences AAC93422 - AAC93449 represent cDNA encoding
 CC human secreted proteins AAB51724 - AAB51777. Sequences AAB51778 -
 CC AAB51825 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences to which they are homologous. The genes and proteins
 CC have activities dependent on the tissues and cells in which they are
 CC expressed. Examples of their activities include immunosuppressive;
 CC antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnerary. The secreted
 CC proteins, polynucleotides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. Oligonucleotide AAC93413 - AAC93421 and peptide AAB51723 are
CC used in the isolation and characterisation of the proteins and
CC polynucleotides of the invention.

XX SQ Sequence 38 AA;

Query Match

Best Local Similarity 1.9%; Score 7; DB 21; Length 38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 VVGNET 102
|||||

DB 29 VVGNET 35

RESULT 6
AAM84417

ID AAM84417 standard; Protein; 61 AA.

XX AC AAM84417;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen SEQ ID NO:12010.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX KM cytostatic; gene therapy; vaccine; metastasis.

OS Homo sapiens.

PN MO200157182-A2.

XX PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227709.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 08-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234224.

PR 25-SEP-2000; 2000US-0234497.

PR 25-SEP-2000; 2000US-0234498.

PR 26-SEP-2000; 2000US-0234998.

PR 27-SEP-2000; 2000US-0235634.

PR 27-SEP-2000; 2000US-0235636.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239335.

PR 13-OCT-2000; 2000US-0239337.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246509.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249255.
PR 17-NOV-2000; 2000US-0249257.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-483426/52.
DR N-PSDB; AAK57198.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 11: SEQ ID NO 12010; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent
CC diagnosis and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 61 AA:

Query Match 1.9%; Score 7; DB 22; Length 61;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 PGFAPRL 299
|||||||

Db 20 PGFAPRL 26
RESULT 7
ID ABB03297 standard; Protein; 75 AA.
XX ABB03297;
XX
XX ABB03297;
XX
XX 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polypeptide SEQ ID NO 1244.
DE
XX
XX Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiatic; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01338.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233397.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234225.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251857.
 PR 08-DEC-2000; 2000US-0251859.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2000US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 PI
 DR WPI: 2001-451937/48.
 DR N-PSDB; AAL34879.
 XX
 XX Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 PS Claim 11: SEQ ID NO 1244; 781pp + Sequence Listing: English.
 XX
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (AB03087-AB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at http://wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 75 AA;
 SQ
 QY 65 VKNGVKG 71
 QY |||||
 Db 56 VKNGVKG 62
 Query Match 1.9%; Score 7; DB 22; Length 75;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 8
 AB012591
 ID AB012591 standard; Protein; 75 AA.
 XX
 AC AB012591;
 XX
 DT 26-FEB-2003 (first entry)
 DT
 XX Novel human musculoskeletal system antigen #211.
 DE

DE Human pancreatic cancer antigen protein sequence SPO ID NO:476.

XX
KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.

XX Homo sapiens.

XX WO20005320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05989.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-579444/54.

XX N-PSDB; AAC98789.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -

XX Claim 11; Page 914; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 76 AA;

Query Match 1.9%; Score 7; DB 21; Length 76;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 210 GADSKGN 216

DB 46 GADSKGN 52

RESULT 10
AAG02242

ID AAG02242 standard; Protein; 83 AA.

XX AAG02242;

XX

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein, SEQ ID NO: 6323.

XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.

XX

PF 21-FEB-2000; 2000EP-0200610.

XX

PR 26-FEB-1999; 99US-0122487.

XX

PA (GEST) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX

DR WPI: 2000-500381/45.

XX

DR N-PSDB; AAC02248.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX

PS Claim 13; SEQ ID 6323; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 83 AA;

Query Match 1.9%; Score 7; DB 21; Length 83;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 346 KALEELA 352

DB 77 KALEELA 83

RESULT 11

AAE10715

ID AAE10715 standard; peptide; 90 AA.

XX

AC AAE10715;

XX

DT 10-DEC-2001 (first entry)

XX

DE Human 4ST3GalIIH domain of stialyltransferase.

XX

KW Human; stialyltransferase; GM3-synthase-specific immunogenic activity;
KW 4ST3GalIIH domain.

XX

OS Homo sapiens.

XX

PN US6280989-B1.

XX

PD 28-AUG-2001.

XX	PF	17-JUN-1999;	99US-0334601.
XX	PR	17-JUN-1999;	99US-0334601.
XX	PA	(KAPI/) KAPITONOV D.	
XX	PI	(YURK/) YU R K.	
XX	PI	Kapitonov D, Yu RK;	
XX	DR	WPI: 2001-579256/65.	
XX	PT	Isolated nucleic acid useful for coding a full-length mammalian GM3 synthase or its biologically active polypeptide fragment, useful for detecting expression of sialyltransferase in whole organs, tissue or cells -	
XX	PS	Example 1; Fig 9A; 81pp; English.	
XX	CC	The present invention relates to an isolated nucleic acid comprising a nucleotide sequence coding for a full-length mammalian GM3 synthase or their biologically active polypeptide fragment. It is also useful as a size marker in nucleic acid or protein electrophoresis or chromatography and for detecting expression of sialyltransferase (ST) in whole organs, tissue or cells. The nucleic acid probes can also be used to detect, quantitate or isolate a mammalian sialyltransferase homologues, as test sample or to identify sialyltransferase reaction products by oligonucleotide probe e.g. in polymerase chain reaction, differential display, in combination with cDNA libraries, expression libraries, etc. The nucleic acid can be modified to make it resistant to cellular enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake into cells. The present sequence is human 45T3GALIH domain of sialyltransferase.	
XX	SQ	Sequence 90 AA;	
XX	Query Match	1.9%; Score 7; DB 22; Length 90;	
XX	Best Local Similarity	100.0%; Pred. No. 72;	
XX	Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY		210 GADSKGN 216 	
DB		76 GADSKGN 82	
RESULT 12			
XX	AAE10716	AAE10716 standard; peptide; 90 AA.	
XX	AC	AAE10716;	
XX	DT	10-DEC-2001. (first entry)	
XX	DE	Mouse 45T3GalIM domain of sialyltransferase.	
XX	RW	Mouse; sialyltransferase; GM3-synthase-specific immunogenic activity; 45T3GalIM domain.	
XX	OS	Mus sp.	
XX	PN	US6280989-B1.	
XX	PD	28-AUG-2001.	
XX	PE	17-JUN-1999; 99US-0334601.	
XX	PR	17-JUN-1999; 99US-0334601.	
XX	PA	(KAPI/) KAPITONOV D.	
XX	PI	(YURK/) YU R K.	
XX	PI	Kapitonov D, Yu RK;	

```

DR      WPI, 2001-579256/65.
XX
XX      Isolated nucleic acid useful for coding a full-length mammalian GM3
PT      synthase or its biologically active polypeptide fragment, useful for
PT      detecting expression of sialyltransferase in whole organs, tissue or
PT      cells -
XX
XX      Example 1; Fig 9A; 81pp; English.
XX
XX      The present invention relates to an isolated nucleic acid comprising a
CC      nucleotide sequence coding for a full-length mammalian GM3 synthase or
CC      their biologically active polypeptide fragment. It is also useful as a
CC      site marker in nucleic acid or protein electrophoresis or chromatography
CC      and for detecting expression of sialyltransferase (ST) in whole organs,
CC      tissue or cells. The nucleic acid probes can also be used to detect,
CC      quantitate or isolate a mammalian sialyltransferase nucleic acid in a
CC      test sample or to identify sialyltransferase homologues, as
CC      oligonucleotide probe e.g. in polymerase chain reaction differential
CC      display, in combination with cDNA libraries, expression libraries, etc.
CC      The nucleic acid can be modified to make it resistant to cellular
CC      enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake
CC      into cells. The present sequence is mouse 4ST3 GalTm domain of
CC      sialyltransferase.
XX
XX      Sequence      90 AA;
SQ
XX
XX      Query Match      1.9%; Score 7; DB 22; Length 90;
XX      Best Local Similarity 100.0%; Pred. No. 72;
XX      Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY      210 GADSKGN 216
XX      |||||||
DB      76 GADSKGN 82

RESULT 13
AAE10717
AAE10717 standard; peptide; 90 AA.
XX
XX      AAE10717;
XX
XX      10-DEC-2001 (first entry)
DT
XX
XX      Pig 4ST3GalIP domain of sialyltransferase.
DE
XX
XX      Pig: sialyltransferase; GM3-synthase-specific immunogenic activity;
KW      4ST3GalIP domain.
XX
XX      Sus scrofa.
OS
XX
XX      US6280989-B1.
PN
XX
XX      28-AUG-2001.
PD
XX
XX      17-JUN-1999; 99US-0334601.
XX
XX      17-JUN-1999; 99US-0334601.
PR
XX
XX      17-JUN-1999; 99US-0334601.
PA      (KAPI/) KAPITONOV D.
PA      (YORK/) YU R K.
XX
XX      Kapitnov D, Yu RK;
XX
XX      WPI, 2001-579256/65.
DR
XX
XX      Isolated nucleic acid useful for coding a full-length mammalian GM3
PT      synthase or its biologically active polypeptide fragment, useful for
PT      detecting expression of sialyltransferase in whole organs, tissue or
PT      cells -
XX
XX      Example 1; Fig 9A; 81pp; English.
XX
XX      The present invention relates to an isolated nucleic acid comprising a

```

CC nucleotide sequence coding for a full-length mammalian GM3 synthase or
 CC their biologically active polypeptide fragment. It is also useful as a
 CC size marker in nucleic acid or protein electrophoresis or chromatography
 CC and for detecting expression of sialyltransferase (ST) in whole organs,
 CC tissue or cells. The nucleic acid probes can also be used to detect,
 CC quantitate or to isolate a mammalian sialyltransferase nucleic acid in a
 CC test sample or to identify sialyltransferase homologues, as
 CC oligonucleotide probe e.g. in polymerase chain reaction differential
 CC display, in combination with cDNA libraries, expression libraries, etc.
 CC The nucleic acid can be modified to make it resistant to cellular
 CC enzymes, oxidation, reduction, nucleases, etc. Or to enhance its uptake
 CC into cells. The present sequence is pig 4ST3 GalIP domain of
 CC sialyltransferase.

CC Sequence 90 AA;

Query Match 1.9%; Score 7; DB 22; Length 90;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 210 GADSKGN 216
 Db 76 GADSKGN 82

RESULT 14

ABG59631
 ID ABG59631 standard; Peptide; 91 AA.

AC ABG59631;

DT 25-FEB-2003 (first entry)

DE Human liver peptide, SEQ ID NO 38279.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
 KW hypercholesterolaemia; coronary heart disease.

OS Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00664.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-488898/53.

Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human adult liver -

Claim 27; SEQ ID NO 38279; 658bp; English.

CC The invention relates to a single exon nucleic acid probe (SENP) (1) for
 CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 13109 defined nucleotide sequences given in the
 CC specification (or complements/fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult
 CC liver. (1) may be used for predicting, measuring and displaying gene
 CC expression in samples derived from human adult liver. The genes
 CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
 CC is associated with coronary heart disease. ABG47348-ABG59930 represent
 CC human liver single exon encoded peptides of the invention.
 CC Note: The sequence information for this patent does not appear in the
 CC printed specification but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 91 AA;

Query Match 1.9%; Score 7; DB 22; Length 91;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 IRLAKEF 112
 Db 43 IRLAKEF 49

RESULT 15

ABB44259
 ID ABB44259 standard; Peptide; 91 AA.

AC ABB44259;

DT 04-FEB-2002 (first entry)

DE Peptide #11765 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

XX WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human foetal liver -

Claim 27; SEQ ID NO 36894; 639bp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 91 AA;

Query Match 1.9%; Score 7; DB 22; Length 91;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 IKLAKKEP 112
| | | | |
Db 43 IKLAKKEP 49

Search completed: July 24, 2003, 21:25:31
Job time : 87 secs

1
2

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:11:05 ; Search time 29 Seconds
(without alignments)
533.992 Million cell updates/sec

Title: US-10-060-848-3
Perfect score: 366
Sequence: 1 MNYSKDAPEVVSFKDAREF.....ALBELAKCVQWMSYKRLK 366

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

1: Issued_Patents-AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
7: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.9	90	US-09-334-601-23	Sequence 23, Appl
2	7	1.9	90	US-09-334-601-24	Sequence 24, Appl
3	7	1.9	90	US-09-334-601-25	Sequence 25, Appl
4	7	1.9	133	US-08-681-812-5	Sequence 5, Appl
5	7	1.9	157	US-09-328-352-5329	Sequence 5329, Ap
6	7	1.9	231	US-09-134-001C-4094	Sequence 4094, Ap
7	7	1.9	238	US-09-107-532A-6232	Sequence 6232, Ap
8	7	1.9	246	US-09-252-991A-29891	Sequence 29891, A
9	7	1.9	260	US-08-495-484-7	Sequence 16, Appl
10	7	1.9	340	US-08-446-875-16	Sequence 2, Appl
11	7	1.9	343	US-08-446-875-2	Sequence 2, Appl
12	7	1.9	343	US-08-102-385G-2	Sequence 2, Appl
13	7	1.9	373	US-09-252-991A-26407	Sequence 26407, A
14	7	1.9	382	US-08-582-740-70	Sequence 9, Appl
15	7	1.9	382	US-09-109-879-70	Sequence 70, Appl
16	7	1.9	387	US-09-107-532A-5675	Sequence 138, App
17	7	1.9	392	US-09-073-009-138	Sequence 68, Appl
18	7	1.9	406	US-08-582-740-68	Sequence 68, Appl
19	7	1.9	406	US-09-109-879-68	Sequence 68, Appl
20	7	1.9	464	US-08-553-619B-9	Sequence 21015, A
21	7	1.9	483	US-09-252-991A-21015	Sequence 21015, A
22	7	1.9	499	US-09-252-991A-31820	Sequence 27, Appl
23	7	1.9	760	US-09-513-057C-27	Sequence 23887, A
24	7	1.9	817	US-09-252-991A-25887	Sequence 23178, A
25	7	1.9	858	US-09-252-991A-23178	Sequence 26016, A
26	7	1.9	896	US-09-252-991A-26816	Sequence 2, Appl
27	7	1.9	1457	US-09-436-874-2	

28	6	1.6	13	4	US-08-817-832B-22	Sequence 22, Appl
29	6	1.6	14	1	US-08-159-340A-25	Sequence 25, Appl
30	6	1.6	28	2	US-08-765-815-7	Sequence 7, Appl
31	6	1.6	28	3	US-08-859-738A-7	Sequence 413, App
32	6	1.6	34	4	US-08-469-260A-413	Sequence 413, App
33	6	1.6	34	4	US-08-488-446-413	Sequence 413, App
34	6	1.6	34	4	US-08-467-344A-413	Sequence 413, App
35	6	1.6	36	4	US-09-149-476-736	Sequence 736, App
36	6	1.6	37	4	US-08-635-886C-281	Sequence 281, App
37	6	1.6	36	3	US-09-019-095A-36	Sequence 422, App
38	6	1.6	56	4	US-08-936-165A-422	Sequence 3, Appl
39	6	1.6	58	1	US-08-215-084A-3	Sequence 3, Appl
40	6	1.6	58	1	US-08-463-212-3	Sequence 3, Appl
41	6	1.6	58	1	US-08-463-211-3	Sequence 3, Appl
42	6	1.6	60	4	US-09-107-532A-5716	Sequence 5716, App
43	6	1.6	72	1	US-08-217-360-6	Sequence 6, Appl
44	6	1.6	72	3	US-08-836-500A-6	Sequence 6, Appl
45	6	1.6	88	4	US-09-489-847-173	Sequence 173, App

ALIGNMENTS

```
RESULT 1
US-09-334-601-23
; Sequence 23, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; TITLE OF INVENTION: NOVEL STALYLTRANSFERASES
; FILE REFERENCE: VCIIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-334-601-23

Query Match
Best Local Similarity 100.0%; Score 7; DB 3; Length 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GADSKGN 216
DB 76 GADSKGN 82

RESULT 2
US-09-334-601-24
; Sequence 24, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; TITLE OF INVENTION: NOVEL STALYLTRANSFERASES
; FILE REFERENCE: VCIIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-09-334-601-24

Query Match
Best Local Similarity 100.0%; Score 7; DB 3; Length 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 210 GADSKGN 216
Db 76 GADSKGN 82

RESULT 3

US-09-334-601-25
; Sequence 25, Application US/09334601
; Patent No. 6280989
; GENERAL INFORMATION:
; APPLICANT: Kapitonov, Dmitri
; APPLICANT: Yu, Robert
; TITLE OF INVENTION: NOVEL STALYTRANSPERASES
; FILE REFERENCE: YCUIP-6
; CURRENT APPLICATION NUMBER: US/09/334,601
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Pig
US-09-334-601-25

Query Match 1.9%; Score 7; DB 3; Length 90;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 210 GADSKGN 216
Db 76 GADSKGN 82

RESULT 4

US-08-681-812-5
; Sequence 5, Application US/08681812
; Patent No. 5763593
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Gadois, Ellen L.
; TITLE OF INVENTION: TBP-Associated Global Negative Regulator
; TITLE OF INVENTION: and Methods of Use Thereof
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,812
; FILING DATE: 29-JUL-1996
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH196-07

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:

LENGTH: 133 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-681-812-5

Query Match 1.9%; Score 7; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 345 IKALEEL 351
Db 57 IKALEEL 63

RESULT 5

US-09-328-352-5329
; Sequence 5329, Application US/09328352
; Patent No. 6562938
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5329
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5329

Query Match 1.9%; Score 7; DB 4; Length 157;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 DVKNGVK 70
Db 57 DVKNGVK 63

RESULT 6

US-09-134-001C-4094
; Sequence 4094, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4094
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4094

Query Match 1.9%; Score 7; DB 4; Length 231;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 AIRLAKE 111
Db 26 AIRLAKE 32

RESULT 7
US-09-107-532A-6232

; Sequence 6232, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...238
; SEQUENCE DESCRIPTION: SEQ ID NO: 6232:
US-09-107-532A-6232

Query Match 1.9%; Score 7; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 LGAAYG 98
DB 15 LLGAYVG 21

RESULT 8
US-09-252-991A-29891
; Sequence 29891, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29891
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29891

Query Match 1.9%; Score 7; DB 4; Length 246;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 PDHAGQL 38
DB 114 PDHAGQL 120

RESULT 9
US-08-495-484-7
; Sequence 7, Application US/08495484
; Patent No. 6329568
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Pang, Sheng-Zhi
; TITLE OF INVENTION: TOMATO SPOTTED WILT VIRUS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/495,484
; FILING DATE: 27-JAN-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10301
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-495-484-7

Query Match 1.9%; Score 7; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 VPKVLKQ 80
DB 45 VPKVLKQ 51

RESULT 10
US-08-446-875-16
; Sequence 16, Application US/08446875
; Patent No. 5658751
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong

APPLICANT: Livingston, Brian Duane
APPLICANT: Gillespie, William
APPLICANT: Kelm, Sorge
APPLICANT: Burlingame, Alma L.
APPLICANT: Medzhradszky, Katalin
TITLE OF INVENTION: Compositions and Methods for the
NUMBER OF SEQUENCES: 16
TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
CORRESPONDENCE ADDRESS:
ADDRESSEE: Poms, Smith, Lande & Rose
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,875
FILING DATE: July 12, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/102,385
FILING DATE: August 4, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
REFERENCE/DOCKET NUMBER: 111-197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-875-16
Query Match 1.9%; Score 7; DB 2; Length 340;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 210 GADSKN 216
DB 290 GADSKN 296
RESULT 11
US-08-446-875-2
Sequence 2, Application US/08446875
Patent No. 5858751
GENERAL INFORMATION:
APPLICANT: Paulson, James C.
APPLICANT: Wen, Xiaohong
APPLICANT: Livingston, Brian Duane
APPLICANT: Gillespie, William
APPLICANT: Kelm, Sorge
APPLICANT: Burlingame, Alma L.
APPLICANT: Medzhradszky, Katalin
TITLE OF INVENTION: Compositions and Methods for the
NUMBER OF SEQUENCES: 16
TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
CORRESPONDENCE ADDRESS:
ADDRESSEE: Poms, Smith, Lande & Rose
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,875
FILING DATE: July 12, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/102,385
FILING DATE: August 4, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421
REFERENCE/DOCKET NUMBER: 111-197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-875-2
Query Match 1.9%; Score 7; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 210 GADSKN 216
DB 293 GADSKN 299
RESULT 12
US-08-102-3856-2
Sequence 2, Application US/08102385G
Patent No. 5962294
GENERAL INFORMATION:
APPLICANT: Paulson, James C.
APPLICANT: Wen, Xiaohong
APPLICANT: Livingston, Brian Duane
APPLICANT: Gillespie, William
APPLICANT: Kelm, Sorge
APPLICANT: Burlingame, Alma L.
APPLICANT: Medzhradszky, Katalin
TITLE OF INVENTION: Compositions and Methods for the
NUMBER OF SEQUENCES: 36
TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
STREET: 2029 Century Park East, 38th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/102,385G
FILING DATE: 04-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/925369
FILING DATE: 04-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Oldenkamp, David J.
REGISTRATION NUMBER: 29,421

REFERENCE/DOCKET NUMBER: 97-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310)788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-102-385G-2

Query Match 1.9%; Score 7; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GADSKGN 216
Db 293 GADSKGN 299

RESULT 13
US-09-252-991A-26407
Sequence 26407, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26407
LENGTH: 373
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26407

Query Match 1.9%; Score 7; DB 4; Length 373;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 ADLLIDA 45
Db 132 ADLLIDA 138

RESULT 14
US-08-582-740-70
Sequence 70, Application US/08582740
Patent No. 6037324
GENERAL INFORMATION:
APPLICANT: Schwender, Charles F.
APPLICANT: Shroff, Hitesh N.
TITLE OF INVENTION: Inhibitors of MADCAM-1-Mediated
TITLE OF INVENTION: Interactions and Methods of Use Therefor
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/582,740
FILING DATE: 04-JAN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: LKS95-12
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-9540
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-582-740-70

Query Match 1.9%; Score 7; DB 3; Length 382;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 LVPGDPE 325
Db 125 LVPGDPE 131

RESULT 15
US-09-109-879-70
Sequence 70, Application US/09109879
Patent No. 6274556
GENERAL INFORMATION:
APPLICANT: Schwender, Charles F.
APPLICANT: Shroff, Hitesh N.
TITLE OF INVENTION: INHIBITORS OF MADCAM-1-MEDIATED
TITLE OF INVENTION: INTERACTIONS AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 89
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02421
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/109,879
FILING DATE: 02-JUL-1998
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/00291
FILING DATE: 03-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/582,740
FILING DATE: 04-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: LKS95-12A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear

; MOLECULE TYPE: protein
US-09-109-879-70

Query Match 1.9%; Score 7; DB 3; Length 382;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 319 LVPGDPE 325
|||
Db 125 LVPGDPE 131

Search completed: July 24, 2003, 21:16:15
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:15:00 ; Search time 65 Seconds
(without alignments)
668.709 Million cell updates/sec

Title: US-10-060-848-3
Perfect score: 366
Sequence: 1 MNSKDAPEFVSPKDAREF.....ALEELATKCDVMEYSKRLK 366

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 451899 seqs, 118759770 residues

Word size : 0

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PC1_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PC105_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.2	176	US-10-156-761-12927	Sequence 12927, A
2	8	2.2	295	US-10-156-761-8626	Sequence 8626, Ap
3	7	1.9	75	US-09-764-877-1244	Sequence 1244, Ap
4	7	1.9	76	US-09-925-297-476	Sequence 476, App
5	7	1.9	89	US-09-798-412-13	Sequence 13, Appl
6	7	1.9	89	US-10-325-917-13	Sequence 13, Appl
7	7	1.9	90	US-09-841-739-7	Sequence 7, Appl
8	7	1.9	90	US-09-767-215-6	Sequence 6, Appl
9	7	1.9	91	US-09-864-761-42428	Sequence 42428, A
10	7	1.9	148	US-08-781-9864-5210	Sequence 5210, Ap
11	7	1.9	200	US-10-156-761-14141	Sequence 14141, A
12	7	1.9	213	US-09-738-626-4676	Sequence 4676, Ap
13	7	1.9	229	US-09-815-242-5786	Sequence 5786, Ap
14	7	1.9	230	US-09-815-242-12784	Sequence 12784, A
15	7	1.9	230	US-09-815-242-13052	Sequence 13052, A

16	7	1.9	240	US-09-978-248-5	Sequence 5, Appl
17	7	1.9	244	US-09-815-242-12364	Sequence 12364, A
18	7	1.9	244	US-09-933-9994-38	Sequence 38, Appl
19	7	1.9	251	US-09-815-242-5409	Sequence 5409, Ap
20	7	1.9	255	US-09-847-6378-9	Sequence 9, Appl
21	7	1.9	318	US-10-156-761-7581	Sequence 7581, Ap
22	7	1.9	337	US-09-800-362-2	Sequence 2, Appl
23	7	1.9	340	US-10-205-823-365	Sequence 365, App
24	7	1.9	345	US-08-875-849C-6	Sequence 6, Appl
25	7	1.9	345	US-09-862-989-6	Sequence 6, Appl
26	7	1.9	366	US-09-938-803-13	Sequence 13, Appl
27	7	1.9	376	US-09-815-242-10220	Sequence 10220, A
28	7	1.9	381	US-10-165-603-20	Sequence 20, Appl
29	7	1.9	382	US-08-875-849C-4	Sequence 4, Appl
30	7	1.9	382	US-09-859-214-70	Sequence 70, Appl
31	7	1.9	382	US-09-862-989-4	Sequence 4, Appl
32	7	1.9	392	US-09-073-009-138	Sequence 138, App
33	7	1.9	392	US-09-793-306-138	Sequence 138, App
34	7	1.9	392	US-10-080-170-610	Sequence 610, App
35	7	1.9	406	US-08-875-849C-2	Sequence 2, Appl
36	7	1.9	406	US-09-859-214-68	Sequence 68, Appl
37	7	1.9	406	US-09-862-989-2	Sequence 2, Appl
38	7	1.9	406	US-10-047-542-68	Sequence 68, Appl
39	7	1.9	413	US-10-156-761-14917	Sequence 14917, A
40	7	1.9	420	US-10-156-761-14741	Sequence 14741, A
41	7	1.9	421	US-09-815-242-11936	Sequence 11936, A
42	7	1.9	475	US-09-769-787-7	Sequence 7, Appl
43	7	1.9	502	US-10-255-817-14	Sequence 14, Appl
44	7	1.9	506	US-10-255-817-12	Sequence 12, Appl
45	7	1.9	580	US-10-156-761-9617	Sequence 9617, Ap

ALIGNMENTS

RESULT 1
US-10-156-761-12927
; Sequence 12927, Application US/10156761
; Publication No. US20030119018A1
GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIDA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12927
LENGTH: 176
TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12927

Query Match 2.2%; Score 8; DB 15; Length 176;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 AGOGLADLL 42
|||||||
DB 155 AGOGLADLL 162

RESULT 2
US-10-156-761-8626

; Sequence 8626, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8626
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8626

Query Match 2.2%; Score 8; DB 15; Length 295;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 VSPRDARE 19
Db 35 VSPRDARE 42

RESULT 3
US-09-764-877-1244
; Sequence 1244, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1244
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (63)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (66)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1244

Query Match 1.9%; Score 7; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 VKNGVGK 71
Db 56 VKNGVGK 62

RESULT 4

US-09-925-297-476
; Sequence 476, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins and antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 476
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (54)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-476

Query Match 1.9%; Score 7; DB 9; Length 76;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GADSKGN 216
Db 46 GADSKGN 52

RESULT 5
US-09-798-412-13
; Sequence 13, Application US/09798412
; Publication No. US20030109428A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; FILE REFERENCE: 07334-327001
; CURRENT APPLICATION NUMBER: US/09/798,412
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 09/728,260
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/685,791
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/513,904
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 09/507,533
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/168,780
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-798-412-13

Query Match 1.9%; Score 7; DB 11; Length 89;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 LADLLLD 44
Db 82 LADLLLD 88

RESULT 6
US-10-325-917-13
; Sequence 13, Application US/10325917
; Publication No. US20030113787A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-327001
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US/09/798,412
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 09/728,260
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/685,791
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/513,904
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 09/507,533
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 60/168,780
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-10-325-917-13

Query Match 1.9%; Score 7; DB 15; Length 89;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 38 LADLLD 44
Db 82 LADLLD 88

RESULT 7
US-09-841-739-7
; Sequence 7, Application US/09841739
; Patent No. US20020034784A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE
; FILE REFERENCE: 07334-329001
; CURRENT APPLICATION NUMBER: US/09/841,739
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 09/697,089
; PRIOR FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: US 60/161,822
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-841-739-7

Query Match 1.9%; Score 7; DB 9; Length 90;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 38 LADLLD 44
|||||||

Db 83 LADLLD 89
RESULT 8
US-09-767-215-6
; Sequence 6, Application US/09767215
; Patent No. US20020081636A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-142001
; CURRENT APPLICATION NUMBER: US/09/767,215
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/181,159
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-767-215-6

Query Match 1.9%; Score 7; DB 9; Length 90;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 38 LADLLD 44
Db 83 LADLLD 89

RESULT 9
US-09-864-761-42428
; Sequence 42428, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42428
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC012578.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
US-09-864-761-42428

Query Match 1.9%; Score 7; DB 9; Length 91;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 IKLAKER 112
Db 43 IKLAKER 49

RESULT 10
US-08-781-986A-5210
Sequence 5210, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 5210:
SEQUENCE CHARACTERISTICS:

LENGTH: 148 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-781-986A-5210

Query Match 1.9%; Score 7; DB 8; Length 148;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 GKVELAD 198
Db 100 GKVELAD 106

RESULT 11
US-10-156-761-14141
Sequence 14141, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14141
LENGTH: 200
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14141

Query Match 1.9%; Score 7; DB 15; Length 200;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 QIADLIL 43
Db 189 QIADLIL 195

RESULT 12
US-09-738-626-4676
Sequence 4676, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIALI, KEIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENO, AKIHITO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16


```

; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4676
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4676

```

```

Query Match          1.9%: Score 7; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      104 LAIKLAK 110
        |||||
Db       164 LAIKLAK 170

```

```

RESULT 13
US-09-815-242-5786
; Sequence 5786, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5786
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5786

```

```

Query Match          1.9%: Score 7; DB 9; Length 229;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      105 AIKLAKE 111
        |||||
Db       26 AIKLAKE 32

```

```

RESULT 14
US-09-815-242-12784
; Sequence 12784, Application US/09815242

```

```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12784
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12784

```

```

Query Match          1.9%: Score 7; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      105 AIKLAKE 111
        |||||
Db       26 AIKLAKE 32

```

```

RESULT 15
US-09-815-242-13052
; Sequence 13052, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13052
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-13052

```

```

Query Match      1.98; Score 7; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      105 AIRLAKE 111
        |||||
Db       26 AIRLAKE 32

```

Search completed: July 24, 2003, 21:23:58
 Job time : 66 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:02:14 ; Search time 83 Seconds
(without alignments)
699.927 Million cell updates/sec

Title: US-10-060-848-3
Perfect score: 1942
Sequence: 1 MNYSKDAPEFVSPKDAREF.....ALEELATKCDVQMFYSYKRLK 366

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/genesqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	785	40.4	446	22	ABB65105
2	504	26.0	385	22	ABB67305
3	472.5	24.3	461	22	ABB60602
4	368	18.9	347	22	AA630483
5	366	18.9	347	22	AA679252
6	306	15.8	357	22	ABB52565
7	248	12.8	332	21	AA707223
8	189.5	9.8	1690	22	ABG25958
9	178.5	9.2	1303	22	ABG12230

10	173.5	8.9	328	22	ABG29197
11	173.5	8.9	1262	22	ABG29003
12	105	5.4	2353	17	AA69393
13	105	5.4	2411	21	ABB23860
14	104.5	5.4	943	19	AAW81745
15	104.5	5.4	943	19	AAW64378
16	104.5	5.4	943	20	AAV39175
17	104.5	5.4	943	20	AAV39032
18	101	5.2	1689	22	ABG16013
19	101	5.2	2417	22	ABG28727
20	99	5.1	940	21	AAV44343
21	98	5.0	883	20	AAV14081
22	98	5.0	940	21	AAW90937
23	98	5.0	940	21	AAV51927
24	98	5.0	940	21	AAV70327
25	97	5.0	655	23	ABB06993
26	94.5	4.9	269	16	ABW73334
27	94.5	4.9	269	22	ABW00043
28	94.5	4.9	382	24	AAE29940
29	93.5	4.8	328	22	ABB65771
30	93.5	4.8	399	22	ABB62871
31	93.5	4.8	443	22	ABB65770
32	92.5	4.8	635	18	AAW25755
33	92.5	4.8	916	24	ABP78856
34	92.5	4.8	1017	20	AAV22427
35	92	4.7	480	20	AAV49108
36	92	4.7	480	23	AAO15075
37	92	4.7	880	24	ABU03490
38	92	4.7	898	20	AAV14082
39	92	4.7	898	21	AAV70326
40	92	4.7	914	20	AAV44344
41	92	4.7	941	20	AAV29796
42	92	4.7	941	21	AAW90938
43	92	4.7	941	21	AAV51928
44	92	4.7	941	21	AAV70328
45	92	4.7	941	21	AAV79202

ALIGNMENTS

RESULT 1	ABB65105	standard; Protein; 446 AA.
ID	ABB65105	
AC	ABB65105;	
DT	26-MAR-2002	(first entry)
DE	Drosophila melanogaster polypeptide SEQ	ID NO 22107.
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.	
KW	Drosophila melanogaster.	
OS	Drosophila melanogaster.	
PN	W0200171042-A2.	
XX	27-SEP-2001.	
XX	23-MAR-2001; 2001WO-US09231.	
XX	23-MAR-2000; 2000US-191637P.	
XX	11-JUL-2000; 2000US-0614150.	
XX	(PEKE) PE CORP NY.	
XX	Venter JC, Adams M, Li PMD, Myers EW;	
XX	WPI: 2001-656860/75.	
XX	N-PSDB: ABL09208.	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more	

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 22107; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 446 AA;
 Query Match 40.4%; Score 785; DB 22; Length 446;
 Best Local Similarity 45.0%; Pred. No. 1.9e-69;
 Matches 163; Conservative 51; Mismatches 142; Indels 6; Gaps 5;
 QY 4 SKAPEVYSPKAREFVVKCMQTVGTSPDHAGQLADLLDADLVGHYSHGLNRLHIYVD 63
 DB 87 SAAPK-LVAVAESRRMIDCFKAVKVPQAHAEQADLLVAADHRGHFSGMNLKMYIN 145
 QY 64 DVK-NGVKGNGVPRVLRKQKGTAVMDGENTLGAIVNGFCDLAIKIKKEGVAMVYTKNS 122
 DB 146 DLAINSTDGAANVRKILKETATATWVDGLNGLGAVVGNICMDLAKRKATYGVGVCNKGK 205
 QY 123 NHYGACOHYTKKIANAGVMSFTNTSPLMPCRSSEIGLGNPLSCVNSEKYGDSFL 182
 DB 206 NHYGMGWYAIKRAMDQGLVGSMTNTSPLMAPTRAKFAALGTNPLSGANA-TNGDKFL 264
 QY 183 DMATFYALGKVELADCKRGTQIPSTGADSKGNPSTDTQVYLHGGLPLGIGEEGSR 242
 DB 265 DMATTAAGVKIETLQRRKG-APLPDGAODPSGEVINDAELGFTGCLMPLGSELTSGY 323
 QY 243 KGTGLSMGELFCGILLAGSSFGKNRVLW--GQSHKADNGOCFAVAIDQCFAPFAPRLQ 300
 DB 324 KGYGLGAMVDILSGVSGANYSQVQRKWTAGADSAADLGQVFAVDPNCFAPFEERMA 383
 QY 301 QPIDEFRNLKPISEKPVLPVGPDERMNTESOKRAGLVYQEOIKALELATKCDVQMF 360
 DB 384 DFNSRLRGATPTPSPKPVLLAGDKKKGMADVDAAGCIQYLENQLKCANLAETLKIKPL 443
 QY 361 SY 362
 DB 444 SF 445
 Db

RESULT 2
 ID ABB67305 standard; Protein; 385 AA.
 AC ABB67305;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 28707.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; AB111408.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 28707; 21pp + Sequence Listing; English.
 CC
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 385 AA;
 Query Match 26.0%; Score 504; DB 22; Length 385;
 Best Local Similarity 32.0%; Pred. No. 1.9e-41;
 Matches 110; Conservative 73; Mismatches 149; Indels 12; Gaps 5;
 QY 16 DAREFVVKCMQTVGTSPDHAGQLADLLDADLVGHYSHGLNRLHIYVDYKNG-VKNGV 74
 DB 38 EAREFIDDCLRGVSPSKYRCISEFLVADVGRNGTGLNRLDYISDLSGHAKEGAE 97
 QY 75 PKVLKQKGTAVMDGENTLGAIVNGFCTDLAIKIKKEGVAMVYTKNSNHYGACOHYTKR 134
 DB 98 PSIISEIVSTAHNGNSALVFVGNFCMDLAVKKAEDSGIGFVAQOSHIDIGMASMPTFR 157
 QY 135 IANAGVMSFTNTSPLMPCRSSEIGLGNPLSCVNSEKYGDSFLDMATFYALGV 194
 DB 158 AAGGGLAGIVMSNAPTPMGPNSKASISGNCFAFCVKGGEY--HFVLDMATSVKIDGAV 215
 QY 195 ELADCKRGTQIPSTGADSKGNPSTDTQVYLHGGLPLGIGETGSKGTGLSMGELF 254
 DB 216 EMWAMNDE-YIPGMANNEGGLSTCFPSLRLRPLLPAGG-----HKGYCLSAVIDIL 268
 QY 255 CGILAGSSFGKNRVLWQSHKADNGOCFAVAIDQCFAPFAPRLQOFLDETRNLKPISE 314
 DB 269 CGVLSGAQVATHTM--DQNPSTLQGVFAIDPEFLPAPFMRPFDFGRIQNSQPAD 326
 QY 315 EKPVLVGPDERMNTESOKRAGLVYQEOIKALELATKCDVQ 358
 DB 327 SEPIRLFGELERHMNYVEDLRALAYPNSLITRYKKEVAERLCYK 370
 Db

RESULT 3
 ID ABB60602 standard; Protein; 461 AA.
 AC ABB60602;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 8598.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 OS

XX WO200171042-A2.
PN 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li FWD, Myers EW;
PI WPI: 2001-656860/75.
XX N-PSDB; ABL04705.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX PS Disclosure; SEQ ID NO 8598; 21pp + Sequence listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB85773-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 461 AA;
SQ
Query Match 24.3%; Score 472.5; DB 22; Length 461;
Best Local Similarity 30.3%; Pred. No. 3.6e-38;
Matches 112; Conservative 66; Mismatches 179; Indels 13; Gaps 6;
QY 2 NYSKDAEEFVNSPDAEEFYKCMQVGTSPDHAGQLADLLDLYGHSHGLRLHIY 61
DB 87 NATKDF-SGLVDVLEAORFVSQVFSAMQVKAASEMDALILADYMGORSMGIHRLPSI 145
QY 62 VDDVKN-GVGNNGVPKYLKOKGTAWNDGEMLGAVNGNCTDLAIKAKFEFVAVMYTK 120
DB 146 AADLLNCTVAGDAPGIVASEKKAIALVDGHNAPGVANFMDLALQKAREVGIGWSAR 205
QY 121 NSNHGACQHYTKRIANAGVMSFTNTSPIMFPCRSSEIGLGNPLSCCVNSEKTGDSF 180
DB 206 SSNCIGFASWYACALQORMIGLCMTNAPLPLRAGIEPLGNERPICAAASG-VHQQF 263
QY 181 LLDMAATTVVALGKVELADCRG-KTQIDPSTWGADSKNGESTDQVVLHGGLPLGIEET 239
DB 264 VADGMAACSVDELELSYCNMGMSKEVPKLVALDRNGKETSTEBALRQRIRAF-----Q 318
QY 240 GSYVGTGLSMWELFCGLLGSSFGKNVR---LWGQSHKADNGOCFAITIOECFAPGA 296
DB 319 PEHGFGLAAVVDILCGMTGARARANOQRRGVYSTENAPANLQVYVAIDPMRCPTFE 378
QY 297 PRLQOFLDETENLPISEKFPVLVPGDPERMNTESQKAGLIVYQEOIKALELATACD 356
DB 379 DRLADFHRLRQANPSPKVGKRPMPVPGDKELQHKMKVNBQGGILTMSSCLTSLVEELATFED 438
QY 357 VQMFSTYRKLK 366
DB 439 IEPLKMRNTK 448

RESULT 4
AAG90483
ID AAG90483 standard; Protein; 347 AA.

XX AAG90483;
AC 26-SEP-2001 (first entry)
XX
XX C glutamicum protein fragment SEQ ID NO: 4237.
DE
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX Ep1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYO) KYOMA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR N-PSDB; AAH65702.
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 17; SEQ ID NO: 4237; 246pp + Sequence listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
SQ
Sequence 347 AA;
SQ
Query Match 18.9%; Score 368; DB 22; Length 347;
Best Local Similarity 31.0%; Pred. No. 6.7e-28;
Matches 110; Conservative 53; Mismatches 164; Indels 28; Gaps 10;
QY 8 PEFVNSPDAEEFYKCMQVGTSPDHAGQLADLLDLYGHSHGLRLHIYDD 64
DB 2 PEVTV--NAOULVLCIDILITKGTVPAAADHLVDSLVQADLWGHGVLRLPWYRR 58
QY 65 VKNG-VKNGNVPKYLKOKGTAWNDGEMLGAVNGNCTDLAIKAKFEFVAVMYTKNSN 123
DB 59 LHSGMTTHAHEVINDGAVLADLGHNGIGOVLAHAKREAVYPAAMFGIGAVSVRNSN 118
QY 124 HYGACQHYTKRIANAGVMSFTNTSPIMFPCRSSEIGLGNPLSCCVNSEKTGDSFLD 183
DB 119 HFGTAMYYTRKAAAGCVSILTTNAPSAMAPGGRKRICTNPMSIAFPGET--ATVVD 176
QY 184 MATTVVALGKVELADCRKKTQIPSTWGADSKNGESTDQVVLHGGLPLGIEETGYSYK 243
DB 177 IANTAVANGKITYHAR-QTNMPDIPETWALTSGAPTTDAEALN-GVVLPMAG-----HK 228

QY	24	GTLSMMGEFFCILLGSSSEFGKKNVRLMGOSHKRAADNGCFVAIDECF--APGAPRIQQ	301
Db	229	GYALISFMMDVLGCVLFGSGHSTKVGHPYPTPPGAGHLFTALDYAARPDQDDALSD	288
QY	302	FLDETRNLKAFISEKFPVLVGPDE---RMTETYSQKAGGLVYQEGQIKALEELA	352
Db	289	LVEGVKSTPKRANTTEIIFYPGESEDDRAHRKNS----AHGISLPEKTWMEIQELA	338
RESULT 5			
ID	AAB79252	standard; Protein; 347 AA.	
AC	AAB79252;		
XX	30-APR-2001	(first entry)	
DE	Corynebacterium glutamicum SMP protein sequence SPO ID NO:20.		
XX			
KW	Corynebacterium glutamicum; carbon metabolism and energy production;		
KW	SMP protein; sugar metabolism and oxidative phosphorylation protein;		
KW	fine chemical production; organic acid; proethnogenic amino acid;		
KW	nonproethnogenic amino acid; purine base; pyrimidine base; nucleoside;		
KW	nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;		
KW	carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;		
XX	diagnosis; Corynebacterium diptheriae; evolutionary study.		
XX	Corynebacterium glutamicum.		
OS	WO200100844-A2.		
PN	04-JAN-2001.		
XX			
XX	23-JUN-2000; 2000WO-IB00943.		
XX			
PR	25-JUN-1999; 99US-0141031.		
PR	08-JUL-1999; 99DE-1031412.		
PR	08-JUL-1999; 99DE-1031413.		
PR	08-JUL-1999; 99DE-1031419.		
PR	08-JUL-1999; 99DE-1031420.		
PR	08-JUL-1999; 99DE-1031424.		
PR	08-JUL-1999; 99DE-1031428.		
PR	08-JUL-1999; 99DE-1031431.		
PR	08-JUL-1999; 99DE-1031433.		
PR	08-JUL-1999; 99DE-1031434.		
PR	08-JUL-1999; 99DE-1031510.		
PR	08-JUL-1999; 99DE-1031562.		
PR	08-JUL-1999; 99DE-1031634.		
PR	09-JUL-1999; 99DE-1032180.		
PR	09-JUL-1999; 99DE-1032227.		
PR	09-JUL-1999; 99DE-1032230.		
PR	09-JUL-1999; 99US-0143208.		
PR	14-JUL-1999; 99DE-1032824.		
PR	14-JUL-1999; 99DE-1032973.		
PR	27-AUG-1999; 99DE-1040765.		
PR	31-AUG-1999; 99US-0151572.		
PR	03-SEP-1999; 99DE-1042076.		
PR	03-SEP-1999; 99DE-1042079.		
PR	03-SEP-1999; 99DE-1042086.		
PR	03-SEP-1999; 99DE-1042087.		
PR	03-SEP-1999; 99DE-1042088.		
PR	03-SEP-1999; 99DE-1042095.		
PR	03-SEP-1999; 99DE-1042123.		
PR	03-SEP-1999; 99DE-1042125.		
XX			
PA	(BADI) BASF AG.		
XX			
PI	Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;		
XX	WPI: 2001-061975/07.		
XX	N-PSDB; AAF71369.		
XX			

PT	New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT	metabolism and oxidative phosphorylation protein for production or
PT	modulation of production of fine chemicals e.g. amino acids,
PT	carbohydrates or enzymes ;
XX	
PS	Claim 20: Page 168-169; 1246pp; English.
XX	
CC	AAE71360 to AAE71750 encode the Corynebacterium glutamicum sugar
CC	metabolism and oxidative phosphorylation (SMP) proteins given in
CC	AAE79243 to AAE 79633 which are involved in carbon metabolism and
CC	energy production. The C. glutamicum SMP gene can be used in vectors
CC	(II) for expression in host cells and production or modulation of
CC	production of fine chemicals, such as, an organic acid, a protinoogenic
CC	or nonprotinoogenic amino acid (preferred), a purine or pyrimidine base,
CC	a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC	acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC	cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
CC	(III) encoded by them are used for diagnosing the presence or activity of
CC	Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
CC	containing them are used to map genomes of organisms related to
CC	C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC	in evolutionary studies, in determining SMP protein regions required
CC	for function, in modulating SMP protein activity, in modulating the
CC	metabolism of sugars, and in modulating high-energy molecule production
CC	in a cell (I.e. ATP, NADPH).
XX	
50	Sequence 347 AA:

Query Match	18.9%	Score 368;	DB 22;	Length 347;
Best Local Similarly	31.0%;	Pred. No. 6.7e-28;		
Matches 110;	Conservative 53;	Mismatches 164;	Indels 28;	Gaps 10

```

QY      8 PEEVYSPKDAAREVYVC---MOTVSPRPHAOQLADLLDDLVGHYSHGIGNRILHIYVD 64
Db      2 PEVTV---NAQOITVLCTDILTKTGVPRAADALVGDLSVQADLMGHPHSGLVRLPWVYR 58

QY      65 VKNG-VKNGCVPEVLKQKGTAMVDGSENLILGAVNCFSTDALAIKLEFEGVAVVYTKSN 122
Db      59 LHSQAMITTAHVHEVDLDGAVLALLOGHNGIGCVLADHARKREAVYPRAMFEGVAVSVRNS 118

QY      124 HVSACOHYTKKILNAGVMGSPFNTSPPLMPCRSSEILGTNPILSCCVNSEKTDSEFLD 183
Db      119 HFGTAVYVTRKAAQCVSILTTNAPSMAAPMGREKRKGTNPMSIAAPFGET--AVYVD 176

QY      184 MATTVYALCKVELADCRKGTQIPSTWGADESKNBPSTDQVYLHGGGLPLPGIIEFTSYK 243
Db      177 IANTAAVARGIYIYAR-QTMMPRIPEWATISBSAPITDPAEAIN-GVYLPMAG-----HK 228

QY      244 GTSLISMAGELFCGILLAGSSFGKNVRLWGSQSHKAADNGOCFAIADQEC--APGEAPRIQQ 301
Db      229 GVAISPMMDVLSGLVLSQHSHTKVGQYDPPPPGAGHILFIALDVAAPRQDPEDDALSD 288

QY      302 FLDETNNLKPISERKVPVLPDGE-----RMNTEYSQKAGGLVYQEGQIKALEELA 352
Db      289 LVEVSVSTPKAONTTEILFYGSESDRAHKNS-----AHGSLSPKTKMWELOELA 338

```

RESULT 6	
ABB52565	
ID	ABB52565 standard; Protein; 357 AA.
XX	
AC	
XX	ABB52565.
XX	
DT	11-FEB-2002 (first entry)
XX	
DE	Escherichia coli polypeptide SEQ ID NO 502.
XX	
KW	Escherichia coli; B2/D+A-; antiinflammatory; antibacterial.
KW	immunosuppressive; extra-intestinal infection; phylogeny; meningitis.
KW	systemic infection; non-diarhoeal infection; septicaemia;
KW	pyelonephritis; antibiotic resistance.
XX	
OS	Escherichia coli.


```

RESULT 8
ABG25958
ID   ABG25958 standard; Protein: 1690 AA.
AC
XX   ABG25958;
XX
XX   18-FEB-2002 (first entry)
XX
XX   Novel human diagnostic protein #25949.
XX
XX   Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX   food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX   Homo sapiens.
XX   WO200175067-A2.
XX   PD   11-OCT-2001.
XX
XX   30-MAR-2001; 2001WO-US08631.
XX
XX   31-MAR-2000; 2000US-0540217.
XX   PR   23-AUG-2000; 2000US-0649167.
XX
XX   (HYSE-) HYSEQ INC.
XX   PA
XX   Drmanac RT, Liu C, Tang YT;
XX   PI
XX   WPI: 2001-639362/73.
XX   DR   N-PSDB; AAS90145.
XX
XX   New isolated polynucleotide and encoded polypeptides, useful in
XX   PT   diagnostics, forensics, gene mapping, identification of mutations
XX   PT   responsible for genetic disorders or other traits and to assess
XX   PT   biodiversity
XX
XX   Claim 20; SEQ ID NO 56317; 103pp; English.
XX
XX   The invention relates to isolated polynucleotide (I) and
XX   CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX   CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX   CC   and gene mapping, and in recombinant production of (II). The
XX   CC   polynucleotides are also used in diagnostics as expressed sequence tags
XX   CC   for identifying expressed genes. (I) is useful in gene therapy techniques
XX   CC   to restore normal activity of (II) or to treat disease states involving
XX   CC   (II). (II) is useful for generating antibodies against it, detecting or
XX   CC   quantitating a polypeptide in tissue, as molecular weight markers and as
XX   CC   a food supplement. (II) and its binding partners are useful in medical
XX   CC   imaging of sites expressing (II). (I) and (II) are useful for treating
XX   CC   disorders involving aberrant protein expression or biological activity.
XX   CC   The polypeptide and polynucleotide sequences have applications in
XX   CC   diagnostics, forensics, gene mapping, identification of mutations
XX   CC   responsible for genetic disorders or other traits to assess biodiversity
XX   CC   and to produce other types of data and products dependent on DNA and
XX   CC   amino acid sequences. ABG00010-ABG30377 represent novel human
XX   CC   diagnostic amino acid sequences of the invention.
XX   CC   Note: The sequence data for this patent did not appear in the printed
XX   CC   specification, but was obtained in electronic format directly from WIPO
XX   CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
XX   SQ   Sequence 1690 AA;
XX
Query Match 9.8%; Score 189.5; DB 22; Length 1690;
Best Local Similarity 29.3%; Pred. No. 5.4e-09;
Matches 46; Conservative 33; Mismatches 75; Indels 3; Gaps 2;
QY 12 VSPKDAREFYVKMOTVGTSPDHAGQLADLLDADLVGHSHGKLNRIHYVDYKNGVG 71
DB 3 ISRETLHQLIENKLCQAGLREHRAATVAELVYADARGISHGAVREYTAERISKG-GT 61
QY 72 NGVPKVLKOKG--TAVVDGENLGAIVNGFCTDLAIKLRKFGVAVWVKNSNHYGACQ 129
DB 62 NRREPERLJETGCSALIHADNMAAGVAAKMGMEHAIKTKAQONGVAVVGISRMGSGAIS 121

```

```

QY 130 HYTKIANAGWGMGFTNTSPIMFPCRSSEIGLCTNP 166
DB 122 YFVGQAARAGFRTGISMCSDPMVYFPGAELIYGTNP 158

RESULT 9
ABG12230
ID   ABG12230 standard; Protein: 1303 AA.
AC
XX   ABG12230;
XX
XX   18-FEB-2002 (first entry)
XX
XX   Novel human diagnostic protein #12221.
XX
XX   Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX   food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX   Homo sapiens.
XX   WO200175067-A2.
XX   PD   11-OCT-2001.
XX
XX   30-MAR-2001; 2001WO-US08631.
XX
XX   31-MAR-2000; 2000US-0540217.
XX   PR   23-AUG-2000; 2000US-0649167.
XX
XX   (HYSE-) HYSEQ INC.
XX   PA
XX   Drmanac RT, Liu C, Tang YT;
XX   PI
XX   WPI: 2001-639362/73.
XX   DR   N-PSDB; AAS76417.
XX
XX   New isolated polynucleotide and encoded polypeptides, useful in
XX   PT   diagnostics, forensics, gene mapping, identification of mutations
XX   PT   responsible for genetic disorders or other traits and to assess
XX   PT   biodiversity
XX
XX   Claim 20; SEQ ID NO 42589; 103pp; English.
XX
XX   The invention relates to isolated polynucleotide (I) and
XX   CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX   CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX   CC   and gene mapping, and in recombinant production of (II). The
XX   CC   polynucleotides are also used in diagnostics as expressed sequence tags
XX   CC   for identifying expressed genes. (I) is useful in gene therapy techniques
XX   CC   to restore normal activity of (II) or to treat disease states involving
XX   CC   (II). (II) is useful for generating antibodies against it, detecting or
XX   CC   quantitating a polypeptide in tissue, as molecular weight markers and as
XX   CC   a food supplement. (II) and its binding partners are useful in medical
XX   CC   imaging of sites expressing (II). (I) and (II) are useful for treating
XX   CC   disorders involving aberrant protein expression or biological activity.
XX   CC   The polypeptide and polynucleotide sequences have applications in
XX   CC   diagnostics, forensics, gene mapping, identification of mutations
XX   CC   responsible for genetic disorders or other traits to assess biodiversity
XX   CC   and to produce other types of data and products dependent on DNA and
XX   CC   amino acid sequences. ABG00010-ABG30377 represent novel human
XX   CC   diagnostic amino acid sequences of the invention.
XX   CC   Note: The sequence data for this patent did not appear in the printed
XX   CC   specification, but was obtained in electronic format directly from WIPO
XX   CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
XX   SQ   Sequence 1303 AA;
XX
Query Match 9.2%; Score 178.5; DB 22; Length 1303;
Best Local Similarity 28.9%; Pred. No. 4.5e-08;
Matches 43; Conservative 29; Mismatches 76; Indels 1; Gaps 1;
QY 29 GTSBDHAGQLADLLDADLVGHSHGKLNRIHYVDYKNG-VKNGVPKVLKOKGTAWY 87

```



```

Db      1122 GVDSSTADACAEEMFARTTESGYSHGVNRPFRFIOLENGDIIIPDAQKRITSLGAIQW 1181
Qy      88 DGENILGAVGNFCDDLAIRAKKEGVAMVYTKSNHNGACOHYTKKTAANAGMGSTFN 147
Db      1182 DAQRISIGMLTKKMMDRALIELAADHGIGLVALRANHMWRGSGYMQAAEKGYIGICWTN 1241
Qy      148 TSPLMFPCRSSEIGTGNPLSCVNSEKT 176
Db      1242 SIAYMPFWGAKCRIGTNPFLVAILPSTGT 1270

RESULT 10
ABG29197
ID      ABG29197 standard; Protein: 328 AA.
XX
XX      ABG29197;
AC
XX      13-FEB-2002 (first entry)
DT
XX
XX      Novel human diagnostic protein #29188.
DE
XX
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX      Homo sapiens.
OS
XX      WO200175067-A2.
XX      11-OCT-2001.
XX      30-MAR-2001; 2001WO-US08631.
XX      31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
XX
XX      (HYSE-) HYSEQ INC.
PA
XX      Drmanac RT, Liu C, Tang YT;
PI
XX
XX      WPI: 2001-639362/73.
DR      N-PSDB; AAS93384.
XX
XX      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity
XX
XX      Claim 20; SEQ ID No 59556; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      a food supplement. (II) and its binding partners are useful in medical
XX      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      disorders involving aberrant protein expression or biological activity.
XX      The polypeptide and polynucleotide sequences have applications in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits to assess biodiversity
XX      and to produce other types of data and products dependent on DNA and
XX      amino acid sequences. ABG00010-ABG30377 represent novel human
XX      diagnostic amino acid sequences of the invention.
XX      Note: The sequence data for this patent did not appear in the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 328 AA;
SQ

```

```

Query Match      8.9%; Score 173.5; DB 22; Length 328;
Best Local Similarity 23.0%; Pred. No. 1.8e-08;
Matches 62; Conservative 39; Mismatches 96; Indels 73; Gaps 7;

Qy      29 GTSPPHAGQLADLDLADLVGHYSHGLNRLHIYVDVNG-VKGNVPRVYLKOKGTAMV 87
Db      20 GVDSSTADACAEEMFARTTESGYSHGVNRPFRFIOLENGDIIIPDAQKRITSLGAIQW 79
Qy      88 DGENILGAVGNFCDDLAIRAKKEGVAMVYTKSNHNGACOHYTKKTAANAGMGSTFN 147
Db      80 DAQRISIGMLTKKMMDRALIELAADHGIGLVALRANHMWRGSGYMQAAEKGYIGICWTN 139
Qy      148 TSPLMFPCRSSEIGTGNPLSCVNS----- 173
Db      140 SIAYMPFWGAKCRIGTNPFLVAILPSTPTTMDMSIKHTLPORAGCTDRKLAMSRQOL 199
Qy      174 -EKTDSPFLDMATTVALGKVE-----LADCRGK-----TQI--P 206
Db      200 LERHGYAF-----NELDQGRPEVTEEEKLFVAACRGREPVTEAERYWSKYMRIRKP 253
Qy      207 STWGADSKGNPSTDTQVYVHLGGGLPLGSI 236
Db      254 KRFTLISGKQPQPKKIVIR-----PLDGL 278

RESULT 11
ABG29003
ID      ABG29003 standard; Protein: 1262 AA.
XX
XX      ABG29003;
AC
XX      18-FEB-2002 (first entry)
DT
XX
XX      Novel human diagnostic protein #28994.
DE
XX
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX      Homo sapiens.
OS
XX      WO200175067-A2.
XX      11-OCT-2001.
XX      30-MAR-2001; 2001WO-US08631.
XX      31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
XX
XX      (HYSE-) HYSEQ INC.
PA
XX      Drmanac RT, Liu C, Tang YT;
PI
XX
XX      WPI: 2001-639362/73.
DR      N-PSDB; AAS93190.
XX
XX      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity
XX
XX      Claim 20; SEQ ID No 59362; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX
XX      Sequence 328 AA;
SQ

```

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_sequences.

XX Sequence 1262 AA:

Query Match 8.9%; Score 173.5; DB 22; Length 1262;

Best Local Similarity 23.0%; Pred. No. 1.4e-07;

Matches 62; Conservative 39; Mismatches 96; Indels 73; Gaps 7;

QY 29 GTSPPHAGQLADLLDADLVGHYSHGLNRLHIYVDVKNQ-VKNGVPLVKOKGCTAWY 87
 DB 20 GVDSEFADACAEKFARTTESGVSHGVNRPREFIQOLENGDIIPDAQPKRITSLGAIEQW 79
 QY 88 DGBNLGAVVGNCTDLAIKAKEFGVAVVYTKNSNHYGACQHYTKKIANAGVMSFTN 147
 DB 80 DQORSTGNTLAKKMPDRALIELADHGIQVLRNANHMWNGSGNGQAAEKGYIGICWTN 139
 QY 148 TSPIMPCRSSEIGLTNPSCCVNS----- 173
 DB 140 STAVMPWPVKECRITNPLIVAIRSTPTVMSMIKHTLPORAACTDKKLMSREAO 199
 QY 174 -EKTGDSFLDMATTVAALGYE-----LADCRK-----TQI--P 206
 DB 200 LEHHGAYF-----NELDLQKREPTVEEKLFAVYVCRGEREPVTEAEVWSKYMTRIKRP 253
 QY 207 STWGADSKGNPSTQVVLHGSGLLPLGCI 236
 DB 254 KRPHITLSGKPKPKKIVIR-----PLPGL 278

RESULT 12

AAR99393 standard; Protein; 2353 AA.

AC AAR99393;

DT 15-JAN-1997 (first entry)

DE Haemophilus adhesion protein HA2.

KW Haemophilus adhesion protein; HA2; hsf protein; vaccine.

OS Haemophilus influenzae type b strain C54.

PN WO9630519-A1.

PD 03-OCT-1996.

PF 22-MAR-1996; 96WO-US04031.

PR 24-MAR-1995; 95US-0409995.

PA (UYSL-) UNIV ST LOUIS.

PA (UNIW) UNIV WASHINGTON.

PI Barenkamp SJ, St Geme JW;

DR WPI; 1996-455364/45.

DR N-PDB; AAT41476.

PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
 XX vaccines against H. influenzae infection.

PS Claim 5; Page 66-73; 120pp; English.

XX Haemophilus adhesion protein HA2 (AAR99393) is associated with the
 CC formation of surface fibrils involved in adhesion to various host
 CC cells; it is also referred to hsf (Haemophilus surface fibrils).
 CC Its amino acid sequence was deduced from a genomic DNA clone
 CC (AAT41476) derived from Haemophilus influenzae type b strain C65.
 CC Large quantities of recombinant HA2 can be produced in transformed
 CC prokaryotic or eukaryotic host cells, for use in vaccines against
 CC H. influenzae infection.

XX Sequence 2353 AA:

Query Match 5.4%; Score 105; DB 17; Length 2353;

Best Local Similarity 23.5%; Pred. No. 2.5;

Matches 98; Conservative 37; Mismatches 162; Indels 120; Gaps 19;

QY 9 EFVYSPKRDAR--EFVYCKMQTVSTPDHAGQLADLLDADLVGHYSHGLNRLHIYVDVYK 66
 DB 1240 EFHAAVKMANAEVEFGKNGATVSAKTIDNNGK-----HTVITDVA 1278
 QY 67 NGVKGNGVPK---VLKOKGTAWVDGENLL-----GAVV--GNF---CTDLAIKL--- 108
 DB 1279 EAKVGDGLEKDTQGTIKLKYDN--TQGNILLTYDATKGSVAKGEFNATYTTDTAQTG 1336
 QY 109 AKEEGVAVVYTKNSNHYGACQHYTKKIANAGVMSFTNTSPIMPCRSSEIGLTNP 168
 DB 1337 ANERGR--VYVKGSCNATATETDKKVAATGVAKAINDAATFVKVENDSATIDSP 1394
 QY 169 CCVNSE-KTGDSEFL-----DMATTVALG-----KELADCRKGTQI 205
 DB 1395 DGANDALKAGDTLTLLAKGNLKYKRDGKNITFPLANDLSYKATVSDKSLGTNGK 1454
 QY 206 PS-----TWGADSK-----GNPSTQVVLHGSGLLPLGIEETGSGYKGTG 250
 DB 1455 TSDTKGKLNPAKDKDTGDDANIHLNGTASTITDTLLNSGARTNIGGNGITDNEKKRA 1514
 QY 251 GELFCGILAGSSFGKAVRLMGOSHKAADNGCFVAIDQCFADFPARLQOFLDETRNLK 310
 DB 1515 DVL-----NAGMWNVR--GVKPAASANN-----QVENID 1539

RESULT 13

AAB23860 standard; Protein; 2411 AA.

AC AAB23860;

DT 17-JAN-2001 (first entry)

DE Haemophilus influenzae adhesin (Hia) protein from type c strain APL.

KW Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;

KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;

KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;

OS Haemophilus influenzae.

PN WO200055191-A2.

PD 21-SEP-2000.

PF 16-MAR-2000; 2000WO-CA00289.

PR 16-MAR-1999; 99US-0268347.

PA (CONN-) CONNAUGHT LAB LTD.

PI Loosmore SM, Yang Y, Klein MH;
 XX WPI: 2000-618897/59.
 DR N-PSDB; AAA92499.
 XX
 PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
 PT use as antigens and vaccines and for treating Hemophilus influenzae
 PT infection -
 XX
 PS Claim 1; Fig 24; 275pp; English.
 XX
 CC The present sequence represents a Haemophilus influenzae adhesin
 CC (Hia) protein from the type c Haemophilus influenzae strain AP1.
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
 CC activities, and can be used in the production of a vaccine. An
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 CC protection against disease caused by Haemophilus strains in a
 CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.
 CC
 XX Sequence 2411 AA;
 SQ
 Query Match 5.4%; Score 105; DB 21; Length 2411;
 Best Local Similarity 23.5%; Pred. No. 2.6;
 Matches 98; Conservative 37; Mismatches 162; Indels 120; Gaps 19;
 QY 9 EFVVSPEKDA--EFVVKCMQTGTSPDHACQLADLIDADLVGHYSHGLNRLHIYVDVK 66
 DB 1298 EFHAAVANKANEVEFGKNGATVSAKTQDNKG-----HTVTIDVA 1336
 QY 67 NGVKGNGVPR-----VLKQKGTAMVQENIL-----GAVY--GNF---CTDLAIKL--- 108
 DB 1337 EAKVGDLEKDTQGIKLVKN--TGNNILITVDATKAGSVAKGEFNAVTTDATTAAGTN 1394
 QY 109 AKEGVAVVYTKSNHGYACQHYTKKIANAGVMSPFTNTPSLMFPERSSEIGLGTNPIS 168
 DB 1395 ANERGR--VYVKGSGNGATATETDKKAYATYGDVAKAINDAATPYKVENDSATIDSPPTD 1452
 QY 169 CCVNSE-KTGDSEFL-----DMATTTVALG-----KVELADCRGKTQI 205
 DB 1453 DGANDALKAGDTLTLKAKRMLKVRKDKNITFALANDLSVKSATVSDKLSIGTNGKNVNI 1512
 QY 206 PS-----TWGADSK-----GNSTDTQVYLHGGGLPLGIEETGSKYKTGIGISM 250
 DB 1513 TSDTKGINFARSKDSTGDDANIHNLGASITLDTLNLGSGATTNLGNGITNEKRAASVK 1572
 QY 251 GELFCGILASSFGKKNRMLGQSHKADNGCCFAVAIDQECFAPGFAPRLQOFLDETRNLK 310
 DB 1573 DVL-----MAGMNR--GVKRPASANN-----QVNIID 1597
 QY 311 PISEEKPV-LVPDPERMNTEYSOKAGLVYQEQIKALBELATKCDVOMFSYKRLK 366
 DB 1598 FVATYDTPDVSDGDKDTTSVTSKDNKG-KTEVKIGAKTSYKIDHNGKLETKELK 1653
 RESULT 14
 ID AAM81745 standard; Protein: 943 AA.
 XX
 AC AAM81745;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide DIF2-1.
 XX

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 XX vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 PA (CORT-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MD;
 PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX
 DR WPI: 1998-261042/23.
 DR N-PSDB; AAV64558.
 PS Example 2; Page 199-202; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
 CC for inducing protective immunity against tuberculosis (TB). This sequence
 CC can be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 CC
 XX Sequence 943 AA;
 SQ
 Query Match 5.4%; Score 104.5; DB 19; Length 943;
 Best Local Similarity 23.8%; Pred. No. 0.7;
 Matches 56; Conservative 25; Mismatches 79; Indels 75; Gaps 13;
 QY 125 YGACQHYTKIANAGVMSPFTNTPSLMFPERSSEIGLGTN-----PLSCVNSEKGT 177
 DB 505 FGNAGDPTNQGPAFNTGNNIGFANTG-----NNNITIGLSGDQGFNLASGNS--GTG 556
 QY 178 DSFLDMATTTVAL-----KVELADCRGKTQIPSTWGAUSKGNPSTDTQVYLHGG--- 228
 DB 557 NSGLFNSGTNNVGIFNAGTGNVGIANSNG-----TGNMGI---GNPGTDNIGILNAGSYNT 608
 QY 229 GLPLG-----GIEETGSKY-----GTGLSMAGELFC 255
 DB 609 GILNAGFNTGFEYNTGSYNTGFEVNGVNTNNGFNVDPTNGSYNPGDPTNNGFENPQVNT 668
 QY 256 GILAGSFGKKNRMLGQSHKADN-GCCFAVAIDQECFAPGFAPRLQOFLDETRNL 309
 DB 669 GAFPTGPFNNCGFLV-----AGDNQGO--IAIDLSVTPP-FIPINQGVAVDINH 714
 RESULT 15
 ID AAM64378 standard; Protein: 943 AA.
 XX
 AC AAM64378;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen DIF2-1.
 XX
 KW Tuberculosis; infection; diagnosis; antigen; DIF2-1.
 XX
 OS Mycobacterium tuberculosis strain Erdman.
 XX
 PN WO9816645-A2.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 20:57:44 ; Search time 29 Seconds

(without alignments)
533.992 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 1942
Sequence: 1 MNSKDAPEFVSPKDAREF.....ALEELATKCDVQMFYKRLK 366

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Postprocessing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PTCDS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	269.5	13.9	332	4	US-09-172-952-22 Sequence 22, Appl
2	257.5	13.3	332	4	US-09-172-952-23 Sequence 23, Appl
3	248	12.8	332	4	US-09-172-952-11 Sequence 11, Appl
4	233.5	12.0	354	4	US-09-252-991A-20132 Sequence 20132, A
5	105	5.4	2353	3	US-09-377-155-33 Sequence 33, Appl
6	105	5.4	2353	3	US-08-913-942-4 Sequence 4, Appl
7	105	5.4	2353	4	US-09-669-974-33 Sequence 33, Appl
8	105	5.4	2411	4	US-09-268-347-36 Sequence 36, Appl
9	104.5	5.4	943	4	US-09-056-556-204 Sequence 204, App
10	104.5	5.4	943	4	US-09-072-596-199 Sequence 199, App
11	104.5	5.4	943	4	US-09-477-135A-131 Sequence 131, App
12	104	5.1	2354	4	US-09-268-347-47 Sequence 47, Appl
13	99	5.1	1912	1	US-08-409-995-4 Sequence 4, Appl
14	99	5.1	1912	3	US-08-685-467-4 Sequence 4, Appl
15	94	4.8	982	4	US-09-252-991A-30580 Sequence 30580, A
16	92.5	4.7	1017	4	US-09-600-776-6 Sequence 6, Appl
17	92	4.7	941	4	US-09-390-134B-31 Sequence 31, Appl
18	90.5	4.7	619	4	US-09-252-991A-26352 Sequence 26352, A
19	88	4.5	665	3	US-09-382-106-2 Sequence 2, Appl
20	88	4.5	742	4	US-09-252-991A-19811 Sequence 19811, A
21	88	4.5	857	2	US-08-779-113-2 Sequence 2, Appl
22	88	4.5	858	2	US-08-583-562B-2 Sequence 2, Appl
23	87	4.5	531	2	US-07-862-588B-7 Sequence 7, Appl
24	87	4.5	795	4	US-09-252-991A-30635 Sequence 30635, A
25	87	4.5	887	1	US-07-867-106-3 Sequence 3, Appl
26	86.5	4.5	492	4	US-09-198-452A-485 Sequence 485, App
27	85.5	4.4	268	1	US-08-431-387-4 Sequence 4, Appl

28	85.5	4.4	1415	4	US-09-252-991A-26438 Sequence 26438, A
29	85	4.4	425	4	US-09-252-991A-20774 Sequence 20774, A
30	85	4.4	501	4	US-09-252-991A-25784 Sequence 25784, A
31	85	4.4	563	4	US-09-134-001C-4800 Sequence 4800, Ap
32	85	4.4	962	4	US-09-071-035-246 Sequence 246, App
33	85	4.4	962	4	US-09-071-035-250 Sequence 250, App
34	85	4.4	962	4	US-09-071-035-254 Sequence 254, App
35	85	4.4	962	4	US-09-071-035-470 Sequence 470, App
36	85	4.4	962	4	US-09-071-035-474 Sequence 474, App
37	85	4.4	962	4	US-09-071-035-478 Sequence 478, App
38	84.5	4.4	1022	3	US-08-772-270A-2 Sequence 2, Appl
39	84	4.3	375	2	US-08-583-562B-10 Sequence 10, Appl
40	84	4.3	375	2	US-08-779-113-10 Sequence 10, Appl
41	84	4.3	430	2	US-08-924-440-2 Sequence 2, Appl
42	84	4.3	532	4	US-09-326-480A-6 Sequence 6, Appl
43	84	4.3	653	4	US-09-298-724-2 Sequence 2, Appl
44	84	4.3	1290	1	US-08-470-350B-2 Sequence 2, Appl
45	83.5	4.3	855	1	US-08-336-343A-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1	US-09-172-952-22	
Sequence 22, Application US/09172952		
Patent No. 6368793		
GENERAL INFORMATION:		
APPLICANT: Hoch, James		
APPLICANT: Dattois, Veronique		
TITLE OF INVENTION: METABOLIC SELECTION METHODS		
FILE REFERENCE: 234/191		
CURRENT APPLICATION NUMBER: US/09/172,952		
CURRENT FILING DATE: 1998-10-14		
NUMBER OF SEQ ID NOS: 33		
SOFTWARE: FastSeq for Windows Version 3.0		
SEQ ID NO 22		
LENGTH: 332		
TYPE: PRT		
ORGANISM: YIAK-EC		
US-09-172-952-22		
Query Match	13.9%	Score 269.5; DB 4; Length 332;
Best Local Similarity	26.1%	Pred. No. 3.9e-20;
Matches	79;	Conservative 64; Mismatches 133; Indels 27; Gaps 8;
QY	29	GTPSPHAGIADLLDADLVGHYSGLNRLHYVDVKNK-VKNGVPRVLKQKGTAVY 87
DB	20	GVDSEFADACAEFARTTESGYSHGVNFPFIOLENGDIIPOAPKRISLAGEW 79
QY	88	DEBNLGAAYGNECTDILAKIKLAKGVAVVTKNSNHGACOHYTKKINAGVGSFTN 147
DB	80	DAORSIGNLTAKKMDRALEADHIGLVALRNHNHWRGSGYQAAEKGYIGICWTN 139
QY	148	TSPIELPPCSSSEIGCTNPLSCCVNSEKTDGSEFLDMATTVALGKVELADCRKQIPS 207
DB	140	STAVNPPWAKKCRISTNPLVAIRSTPI---TWDMSSMSFSGYLEVNRLAGR QLFV 195
QY	208	TWGADESKNPSTDTOVYVHLGGLLPLGIEETGSYKGLSMNGELFCGILAGSGFKNV 267
DB	196	DGDFDEGNLTKEPVEYIEKRNRLPM-----GYWKGSGMSIVLDMIAVTL--SDGASV 246
QY	268	RIMGOSHKADN-----GQCVAI--DDECAFGAPAPLOQFDETRKLRISEKPYL 319
DB	247	-----AEVYQDSNDEYGISQIFIALEVDKLDGPFADAKLQIMYVVSABRADNOAIR 301
QY	320	VPG 322
DB	302	LPG 304
RESULT 2	US-09-172-952-23	


```

APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-6105-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

Query Match          5.4%; Score 105; DB 3; Length 2353;
Best local similarity 23.5%; Pred. 0.27;
Matches 98; Conservative 37; Mismatches 162; Indels 120; Gaps 19;

          9 EFVVPKPDAR--EFVVKCMQTVGTSPHAGQLADLLDADLVGHYSHGLNRLIYVDVK 66
          1240 EFHAAYVKNANEVEFEVKGKNGATVSAKTDNNK-----HTVTIDVA 1278
          67 NGVGNQNVPR-----VLKOKGSTAVDDENL-----GAVY--GNF---CTDLAIKL--- 108
          1279 EAKVGGDLKEDTDGKIKLKVDN--TDGNNLLTVDATKGASAKGEFNAVTIDATTAQGTN 1336
          109 AKERGVAVVYTKSNHNGACOHYTKKIANAGVMGMSFTNTSPLMPCRSSEIGLGTNPLS 168
          1337 ANERGR--VYVKGSGNATATETDCKKAVATGVDAKAINDATPEVKVENDSATIDDSPTD 1394
          QY 169 CCVNSE-KTGDSPFL-----DMATTVALG-----KVELADCRGKIQT 205
          DB 1395 DGANDALKADGDTLTLLKAGKMLKVRDKDKNTTFALANDLSYKSAFVSDKLSGTGNKRYNI 1454
          QY 206 PS-----TWGADSK-----GNPSTDQVVLHGGGLPLLGIEETGSGYKGTGLSM 250
          DB 1455 TSDPTKGLNFAKDSKSTGTDANHLNGINSTLTDLTLNNGATNTLGGNGITDEKKRAASVK 1514
          QY 251 GELFCGTLAAGSFRKKNVRLMGQSHKAADNGCFVAIDQECFAPGFAPRLQOFLDETRNLK 310
          DB 1515 DVL-----MGWNVPR--GVKRPASANN-----QVENID 1539
          QY 311 PISEKRPV-LVPGDPEPMNTEYSQAGGLVYQEGQIKALEBTLATKCDVOMFSYKRLK 366
          DB 1540 FVAATYDVIDVSGDKDITTSYTVESKDNG-KTEVKIGAKTSYIVDNHGKLFPTGKELK 1555

RESULT 7
US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 6331173
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669, 974
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353

```

TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match 5.48; Score 105; DB 4; Length 2353;
Best Local Similarity 23.58; Pred. No. 0.27;
Matches 98; Conservative 37; Mismatches 162; Indels 120; Gaps 19;

QY EFVVSPPDAR--EFVVKCMQTVGTPDHAGOLADLLDADLVGHYSHGLNRLHIYVDYK 66
DB EFHAAVKNANNEVEFGKNGATVSAKTIDNNGK-----HVTYIDVA 1278
QY NGVKGNGVPR---VLKQKGTAVNDGENLL-----GAVY--GNF---CTDLAIKL--- 108
DB EAKVGDGLERDIDGKIKIKLYDN--TDGNNLLTVATKASVAKGEFNAAVTDTAAGTN 1336
QY 109 AKEGVAMVYTKNSNHGACQHYTKKIANAGVMGSMSTNTPLMFPCRSSEIGLGNPLS 168
DB 1337 ANERGR--VVVKGSGNATATETDKKKAIVYGDVAKAINDATFVKVENDSATIDSPD 1394
QY 169 CCVNSE-KTGDSEFL-----DMATTVALG-----KVELADCRGKTQI 205
DB 1395 DGANDALKAGDTLLKAGKNLKYKRDGNITFALANDLSVKSATVSDKLSLGTGNKNYNI 1454
QY 206 PS-----TWGADSK-----GNPSTDVQVVLHGGGLPLGIEETGSGYKGTGLSM 250
DB 1455 TSDTKGLNFAKDSKTGDANIHLNGIASTLTDTLLNSGATTNLGNGITDNEKKRAASVK 1514
QY 251 GELFCGILAGSSFGKNVRLMGOSHKAADNGCCFVAIDQECFAPFARLQOFLDETRNLK 310
DB 1515 DVL-----NAGMNVN--GVKPASANN-----QVENID 1539
QY 311 PISEEKV-LVPPDPERMNTREYSQKAGLVYQGOIKALEELATKCDVOMFSYKRLK 366
DB 1540 FVATYDTVDPSGDKDTTSVYVESKDNG-KREYKIGAKTSVINDHNGKLFTGELK 1595

RESULT 8

US-09-268-347-36
Sequence 36, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 36
LENGTH: 2411
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-268-347-36

Query Match 5.48; Score 105; DB 4; Length 2411;
Best Local Similarity 23.58; Pred. No. 0.28;
Matches 98; Conservative 37; Mismatches 162; Indels 120; Gaps 19;

QY EFVVSPPDAR--EFVVKCMQTVGTPDHAGOLADLLDADLVGHYSHGLNRLHIYVDYK 66
DB 1298 EFHAAVKNANNEVEFGKNGATVSAKTIDNNGK-----HVTYIDVA 1336
QY 67 NGVKGNGVPR---VLKQKGTAVNDGENLL-----GAVY--GNF---CTDLAIKL--- 108
DB 1337 EAKVGDGLERDIDGKIKIKLYDN--TDGNNLLTVATKASVAKGEFNAAVTDTAAGTN 1394
QY 109 AKEGVAMVYTKNSNHGACQHYTKKIANAGVMGSMSTNTPLMFPCRSSEIGLGNPLS 168
DB 1395 ANERGR--VVVKGSGNATATETDKKKAIVYGDVAKAINDATFVKVENDSATIDSPD 1452
QY 169 CCVNSE-KTGDSEFL-----DMATTVALG-----KVELADCRGKTQI 205

DB 1453 DGANDALKAGDTLLKAGKNLKYKRDGNITFALANDLSVKSATVSDKLSLGTGNKNYNI 1512
QY 206 PS-----TWGADSK-----GNPSTDVQVVLHGGGLPLGIEETGSGYKGTGLSM 250
DB 1513 TSDTKGLNFAKDSKTGDANIHLNGIASTLTDTLLNSGATTNLGNGITDNEKKRAASVK 1572
QY 251 GELFCGILAGSSFGKNVRLMGOSHKAADNGCCFVAIDQECFAPFARLQOFLDETRNLK 310
DB 1573 DVL-----NAGMNVN--GVKPASANN-----QVENID 1597
QY 311 PISEEKV-LVPPDPERMNTREYSQKAGLVYQGOIKALEELATKCDVOMFSYKRLK 366
DB 1598 FVATYDTVDPSGDKDTTSVYVESKDNG-KREYKIGAKTSVINDHNGKLFTGELK 1653

RESULT 9

US-09-056-556-204
Sequence 204, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-204

Query Match 5.48; Score 104.5; DB 4; Length 943;
Best Local Similarity 23.88; Pred. No. 0.072;
Matches 56; Conservative 25; Mismatches 79; Indels 75; Gaps 13;

QY 125 YGACQHYTKKIANAGVMGSMSTNTPLMFPCRSSEIGLGN-----PLSCVNSEKGT 177
DB 505 FGNAGDPNCGFANNGNINIFANNG-----NNNIGIGLGNQOGFNIAAGMNS-GTG 556
QY 178 DSPFLDMATTVAL-----KVELADCRGKTQIPSTWGAASKGNPSTDTQVYVLAG- 228
DB 557 NSGLEPNSGTNNVGFNAGTGVNGLANS- - - - -TGNGI--GNPGTDNTGILNAGSYNT 608
QY 229 GLPLIG-----GIEETGSYK-----GTGLSMGELFC 255
DB 609 GILNAGDPNCGFNTGYSYNTNGFVNGTNTNGFVNGDTNTNGSYNPGDTNTGFPNPGVNT 668
QY 256 GILAGSSFGKNVRLMGOSHKAADN-GCCFVAIDQECFAPFARLQOFLDETRNL 309


```

Db 1279 EAKVGDLEKDTGKIKLKYDN--TDGNNLLTVDATGASVANGEFNAVTTDAITTAOGTN 1336
QY 109 AKREGVAVVYTKNSNHGACQHYTKKIANAGVMGMSFTNTSPLMPCRSSEIGIGTPLS 168
Db 1337 ANERGR--VYVKGSGNGATATETDKKVAATYGDVAKAINDATAFKVENDSATITDDEPTD 1394
QY 169 CCVNSE-KTGDSFLL-----DMATTTVALG-----KVELADCRGKTQI 205
Db 1395 DGANDALKADDTLTLKAGKNILKVRDGNITTFALANDLSVKSATVSDKLSLGTGNKNVNI 1454
QY 206 PS-----TWGADSK-----GNPSTDTQVVLHGGGLPLGIEETSGYKGTGLSM 250
Db 1455 TSDTKGLKFAKDSKTGDDANIHLNGIASTLTDLLNSGATTNLGGNGITDNEKKRAASVK 1514
QY 251 GELFCGILAGSSFGKRVRLMGOSHKAADNGOCFAVAIDQECFAPGAPRLQOFLDETRNLK 310
Db 1515 DVL-----MAGMNVN--GVKPAASANN-----QVENID 1539
QY 311 PISEKEVY-LVPGDPERMNTSEYOKAGGLVYOGQITALEBLATKCDVQMFYSYKRLK 366
Db 1540 FVATYDVTDFVSGDKDTTSTVTSKSDNG-KRTEVKIGAKTSVIKDHNGKLTGTKEK 1595

```

RESULT 13

```

US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-4

```

Query Match

Best Local Similarity 5.1%; Score 99; DB 1; Length 1912;

Matches 97; Conservative 37; Mismatches 163; Indels 120; Gaps 19;

```

QY 9 EFVVSPEKDA--EFVVKCQVGTSPDHAGQLADLLDADLVGHYSHGLNRHLHYDDVK 66
Db 1239 EFHAAVYANANEVFGVNGATVSAKTDNNGK-----HVTVIDVA 1277
QY 67 NGVKGNGVPR-----VLKQKGTAMVGENLL-----GAVV--GNF---CTDLAIKL--- 108

```

```

Db 1278 EAKVGDLEKDTGKIKLKYDN--TDGNNLLTVDATGASVANGEFNAVTTDAITTAOGTN 1335
QY 109 AKREGVAVVYTKNSNHGACQHYTKKIANAGVMGMSFTNTSPLMPCRSSEIGIGTPLS 168
Db 1336 ANERGR--VYVKGSGNGATATETDKKVAATYGDVAKAINDATAFKVENDSATITDDEPTD 1393
QY 169 CCVNSE-KTGDSFLL-----DMATTTVALG-----KVELADCRGKTQI 205
Db 1394 DGANDALKADDTLTLKAGKNILKVRDGNITTFALANDLSVKSATVSDKLSLGTGNKNVNI 1453
QY 206 PS-----TWGADSK-----GNPSTDTQVVLHGGGLPLGIEETSGYKGTGLSM 250
Db 1454 TSDTKGLNFAKDSKTGDDANIHLNGIASTLTDLLNSGATTNLGGNGITDNEKKRAASVK 1513
QY 251 GELFCGILAGSSFGKRVRLMGOSHKAADNGOCFAVAIDQECFAPGAPRLQOFLDETRNLK 310
Db 1514 DVL-----MAGMNVN--GVKPAASANN-----QVENID 1538
QY 311 PISEKEVY-LVPGDPERMNTSEYOKAGGLVYOGQITALEBLATKCDVQMFYSYKRLK 366
Db 1539 FVATYDVTDFVSGDKDTTSTVTSKSDNG-KRTEVKIGAKTSVIKDHNGKLTGTKEK 1594

```

RESULT 14

```

US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

```

Query Match

Best Local Similarity 5.1%; Score 99; DB 3; Length 1912;

Matches 97; Conservative 37; Mismatches 163; Indels 120; Gaps 19;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:01:34 ; Search time 70 Seconds

(without alignments)
620.944 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 1942

Sequence: 1 MNYSKDAPEFVVSFKDAREF.....ALEELATKCDVQMFSTKRLK 366

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

451899

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications -AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	368	18.9	347	10	US-09-738-626-4237
2	269.5	13.9	332	11	US-09-557-796-22
3	257.5	13.3	332	11	US-09-557-796-23
4	248	12.8	332	11	US-09-557-796-11
5	105	5.4	2353	10	US-09-797-862-33
6	104.5	5.4	943	10	US-09-966-634-131
7	104.5	5.4	943	11	US-09-997-182-131
8	104.5	5.4	943	11	US-09-997-181-131
9	98	5.0	929	10	US-09-793-139-4
10	98	5.0	940	9	US-09-818-879-4
11	98	5.0	940	9	US-09-211-755B-4
12	94.5	4.9	269	11	US-09-776-334A-7
13	94.5	4.9	382	11	US-09-813-408-7
14	92.5	4.8	1017	10	US-09-965-830-6
15	92	4.7	898	9	US-09-818-879-2

16	92	4.7	898	9	US-09-211-755B-2	Sequence 2, Appl1
17	92	4.7	898	10	US-09-793-139-2	Sequence 2, Appl1
18	92	4.7	941	9	US-09-818-879-47	Sequence 47, Appl1
19	92	4.7	941	9	US-09-211-755B-47	Sequence 47, Appl1
20	92	4.7	941	10	US-09-793-139-47	Sequence 47, Appl1
21	92	4.7	941	15	US-10-300-616-31	Sequence 31, Appl1
22	92	4.7	941	15	US-10-225-567A-436	Sequence 436, App
23	90.5	4.7	580	15	US-10-156-761-888A	Sequence 888A, Ap
24	90.5	4.7	660	15	US-10-101-464A-808	Sequence 808, App
25	90	4.6	871	15	US-10-195-144-45	Sequence 45, Appl1
26	88	4.5	255	10	US-09-847-637B-9	Sequence 9, Appl1
27	88	4.5	665	9	US-09-815-242-11792	Sequence 11792, A
28	87.5	4.5	777	11	US-09-733-643-15	Sequence 15, Appl1
29	85.5	4.4	424	15	US-10-156-761-9629	Sequence 9629, Ap
30	85.5	4.4	569	15	US-10-214-524-30	Sequence 30, Appl1
31	85.5	4.4	638	15	US-10-156-761-12352	Sequence 12352, A
32	85.5	4.4	978	14	US-10-007-693-65	Sequence 65, Appl1
33	85.5	4.4	1116	10	US-09-977-577-10	Sequence 10, Appl1
34	85.5	4.4	1149	10	US-09-977-577-11	Sequence 11, Appl1
35	85.5	4.4	1151	10	US-09-977-577-13	Sequence 13, Appl1
36	85.5	4.4	1156	10	US-09-977-577-12	Sequence 12, Appl1
37	85	4.4	370	10	US-09-931-457A-60	Sequence 60, Appl1
38	84.5	4.4	1032	11	US-09-733-643-16	Sequence 16, Appl1
39	84.5	4.4	1907	10	US-09-832-292-39	Sequence 39, Appl1
40	84	4.3	406	15	US-10-286-264-46	Sequence 46, Appl1
41	84	4.3	406	16	US-10-278-536-158	Sequence 158, App
42	83.5	4.3	520	10	US-09-771-161A-105	Sequence 105, App
43	83.5	4.3	855	10	US-09-771-161A-196	Sequence 196, Appl1
44	83.5	4.3	855	11	US-09-355-815-6	Sequence 6, Appl1
45	83.5	4.3	926	14	US-10-023-437-57	Sequence 57, Appl1

ALIGNMENTS

RESULT 1

US-09-738-626-4237

Sequence 4237, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKOKO

APPLICANT: SENO, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 4237

LENGTH: 347

TYPE: PRT

ORGANISM: Corynebacterium glutamicum

US-09-738-626-4237

Query Match 18.9%; Score 368; DB 10; Length 347;

Best Local Similarity 31.0%; Pred. No. 1.3e-29;

Matches 110; Conservative 53; Mismatches 164; Indels 28; Gaps 10;

QY 8 PERVSPKAREFVVC--MOTVGTSPHAGQLDLLDADLVGHSHGLRLHIYVD 64


```

Db      40 GYSHGVNRPRFQIQLDN---GDILPEQPOQRVTTLAGIEQWMDQRSIGNLAKKMDR 96
      105 AITAKKEGVAVVYTKNSNHGACOHYTKKIANAGVGSFTNTSPLMFPCRSSEIGCT 164
      97 ATELSADHGIGIALVARNANHMWRGSGYQWAAEKGYIGICWTNSIAVAMPKACRIST 156
      165 NPLSCVNSEKSGDSFLDMATTVALGKVELADRGKTQIPSTMGASKGNPSTDTQV 224
      157 NPLVATPSTP---TWVDMSSMPSYGLAEVNRLAGR-ELFVDDGFGDDGRLTKPEGTI 212
      225 LHGGILPLGGIETGYSYKGTGLSMNGELFCGILA-GSSFG---KNVRLMGOSKKAADN 279
      213 EKNRRLPM-----GYKMGSLSTVLMIAATLNGSSVAEVQENSDSEYGV----- 260
      280 GOCFVAIDQECFAPFA--PRLQPLDETNRINKPISEKPVLPFC-----DPERNN 328
      261 -QIFTAIEVDKLIDATRDAKLQRIIMDEFTTAERADENVAVRLPGHEFTRLDENRRN 317

```

RESULT 5

```

US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

```

Query Match 5.4%; Score 105; DB 10; Length 2353;

Best Local Similarity 23.5%; Pred. No. 0.61;

Matches 98; Conservative 37; Mismatches 162; Indels 120; Gaps 19;

```

      9 EFVVPKRDAR--EFVVKCMQTVGTPSPDHAGOLADLLDLADLVGHYSHGLRLHYVDVK 66
      1240 EFHAIVKANEVEYFGKATYSAKTDDNGK-----HYVTIDVA 1278
      67 NGVKGNGVVK---VLKQKGTAVWGENLL-----GAVV--GNF---CTDLAIKL--- 108
      1279 EAKVVDGDEKTDGKIKLKVDN--TDGNLLTVDATKGSVAKGEFNNAVTTATTAQGTN 1336
      109 AKKEGVAVVYTKNSNHGACOHYTKKIANAGVGSFTNTSPLMFPCRSSEIGCTNPIS 168
      1337 ANERK--VYVKSNGATATETDKRVAIVGDAVAINDAAAFVENDSDATIDSPDK 1394
      169 CCVNSE-KTGDSEFL-----DMAATTVALG-----KVELADCRGKTQI 205
      1395 DGANALAKAGDTLTILAKAKNLKVKRDKNITFALANDLSYKSATYSDKISLGTNGKVAI 1454
      206 PS-----TWGADSK-----GNPSTDTQVVLHGGGLPLGGIETGYSYKGTGLSM 250
      1455 TSDTKGLNFAKDSKTGDANIHNLGIASTLTDTLLNSGATWLNCGITDNKKRAASVK 1514
      251 GELFGIILAGSFGKNVRLMGOSHKAAADNGOCFVAIDQECFAPRRLQOFLDETRNLK 310
      1515 DVL-----NAGNNVR--GVKPASANN-----GVEND 1539
      311 PISEKPV-LVPGDEPRMNTESQKAGLIVYQGOIKALELATAKCDVQMFYKRLK 366

```

```

Db      1540 FVATDTDFVFGSDKDTTSVYVESKNG-KRTEVRIKATSVYIKOHNGNLFTGKEELK 1595

```

RESULT 6

```

US-09-996-634-131
; Sequence 131, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

```

Query Match 5.4%; Score 104.5; DB 10; Length 943;

Best Local Similarity 23.8%; Pred. No. 0.16;

Matches 56; Conservative 25; Mismatches 79; Indels 75; Gaps 13;

```

      125 YGACHYTKKIANAGVGSFTNTSPLMFPCRSSEIGCTN-----PLSCVNSEKGT 177
      612 FGNAGDFNGCFANTGNNNGFANTG-----NNNIGILSDNDQNGFIAAGWS-GRG 663
      178 DSFLDMATTVAL-----GKVELADCRGKTQIPSTWADSKGNPSTDTQVVLHGG---- 228
      664 NSGLFNSTGNNNGIFNAGTGNVGIANSNG-----TGNMGI---GNPCTDNTGILNAGSYNT 715
      229 GLPLG---GIEFTGYSK-----GTGSLMNGELFC 255
      716 GILNAGDFNTGFFYNGSYNTGFFNVGNTNTGNNVGDNTGTYNPGDNTTGFNPGNVNT 775
      776 GAFDIDGFNNGFLV-----AGDNGQ--IAIDLSTVTP-FIPINDQMYIDVHNV 821

```

RESULT 7

```

US-09-997-182-131
; Sequence 131, Application US/09997182
; Patent No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0

```

SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match 5.4%; Score 104.5; DB 11; Length 943;
Best Local Similarity 23.8%; Pred. No. 0.16;
Matches 56; Conservative 25; Mismatches 79; Indels 75; Gaps 13;

Y 125 YGACQHTKRIANAGVMSFTNTSPLMFPCRSSEIGLGTN-----PLSCCVNSKGTG 177
D 612 FGNAAGDPNQGFAVNTGNNNIGFANTG-----NNNIGIGLSCDNGQGFNIAGSNTS-GTG 663
Y 178 DSFLDMATTVAL-----GKVELADCRGKTQIPSTWAGDSKGNPSTDTQVVLHGG----- 228
D 664 NSGLFNSGTNNVGFNAGTGNVGIANGS-----TGNNGI---GNPQDNTNGILNAGSYNT 715
Y 229 GLPLG-----GIETGSKY-----GTGLSMGGLFC 255
D 716 GILNAGDPNTGFYNTGNTGFFNVGNTNGNFVGDNTNGSYNPGDNTNGFFNPGVNT 775
Y 256 GILASSEFGKAVRLMGSHKADN-GOCFVAIDGECFAPGAPRLQOFLDETRNL 309
D 776 GAFDTGDFNNGFLV-----AGDNQGO--IAIDLSTVTP-FIPINEQAVIDVHNV 821

RESULT 8

US-09-997-181-131
Sequence 131, Application US/09997181
Publication No. US20030049269A1
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 61257
CURRENT APPLICATION NUMBER: US/09/997,181
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match 5.4%; Score 104.5; DB 11; Length 943;
Best Local Similarity 23.8%; Pred. No. 0.16;
Matches 56; Conservative 25; Mismatches 79; Indels 75; Gaps 13;

Y 125 YGACQHTKRIANAGVMSFTNTSPLMFPCRSSEIGLGTN-----PLSCCVNSKGTG 177
D 612 FGNAAGDPNQGFAVNTGNNNIGFANTG-----NNNIGIGLSCDNGQGFNIAGSNTS-GTG 663
Y 178 DSFLDMATTVAL-----GKVELADCRGKTQIPSTWAGDSKGNPSTDTQVVLHGG----- 228
D 664 NSGLFNSGTNNVGFNAGTGNVGIANGS-----TGNNGI---GNPQDNTNGILNAGSYNT 715
Y 229 GLPLG-----GIETGSKY-----GTGLSMGGLFC 255
D 716 GILNAGDPNTGFYNTGNTGFFNVGNTNGNFVGDNTNGSYNPGDNTNGFFNPGVNT 775
Y 256 GILASSEFGKAVRLMGSHKADN-GOCFVAIDGECFAPGAPRLQOFLDETRNL 309
D 776 GAFDTGDFNNGFLV-----AGDNQGO--IAIDLSTVTP-FIPINEQAVIDVHNV 821

RESULT 9

US-09-793-139-4
Sequence 4, Application US/09793139
Patent No. US20020156265A1
GENERAL INFORMATION:
APPLICANT: Jones, Kenneth A
TITLE OF INVENTION: DNA Encoding A GABA BR2 Polypeptide And Uses Thereof
FILE REFERENCE: 54002epctus
CURRENT APPLICATION NUMBER: US/09/793,139
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 929
TYPE: PRT
ORGANISM: Rattus Sp.
US-09-793-139-4

Query Match 5.0%; Score 98; DB 10; Length 929;
Best Local Similarity 19.5%; Pred. No. 0.76;
Matches 53; Conservative 35; Mismatches 80; Indels 104; Gaps 12;

Y 123 NHR-----GACQHTKRIANA-----GVGMSFTNTSPLMFPCRSSEIGLGTNPLSCCVN-- 172
D 125 NHRVFGVCPVSVTSIAESLQGNLVLQSFATTPVLADKKKYPFFRTVPSDNAAVPA 184
Y 173 -----SEKTDSPFLDMATTVALGK-VELADCRGKTQIPST 208
D 185 ILKILKHFRRRGTLTQDVORSEVND-----LTGLVLYGDEIHSPTESPSNDPCT 237
Y 209 WGSADSKNPSTDTQVVLHGGGLPLGIEFTGSKYGTGLSMGGLFCGIIAGSFGKAVR 268
D 238 SVKTKKN---DVRIL-----LQGFQD-----MMAKVFCCAFEESEFGSKYQ 277
Y 269 -----LMGSHKADNGOC-----FVAIDGECFAPGAPRLQOFLDETR 307
D 278 WIIPGWEPAMVEQVHVEANSRCLRRSLAAMEGYIGVDFE----- 319
Y 308 NLRPSEKPYLPG-DPERMNEYSSOKAGL 338
D 320 ---PLSKQIKTTISGKTPQOTEREYNSKRGV 348

RESULT 10

US-09-818-879-4
Sequence 4, Application US/09818879
Patent No. US20010023289A1
GENERAL INFORMATION:
APPLICANT: Jones, Kenneth
APPLICANT: Laz, Thomas
APPLICANT: Borowsky, Beth
TITLE OF INVENTION: DNA encoding a GABABR2 polypeptide and uses thereof
FILE REFERENCE: 1795/54002DA
CURRENT APPLICATION NUMBER: US/09/818,879
CURRENT FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: US 09/211,755
PRIOR FILING DATE: 1998-12-15
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 940
TYPE: PRT
ORGANISM: Rattus sp
US-09-818-879-4

Query Match 5.0%; Score 98; DB 9; Length 940;
Best Local Similarity 19.5%; Pred. No. 0.77;
Matches 53; Conservative 35; Mismatches 80; Indels 104; Gaps 12;

Y 123 NHR-----GACQHTKRIANA-----GVGMSFTNTSPLMFPCRSSEIGLGTNPLSCCVN-- 172
D 125 NHRVFGVCPVSVTSIAESLQGNLVLQSFATTPVLADKKKYPFFRTVPSDNAAVPA 184


```

Db      125 NMLAVGVCPSVTSLIAESLOGMNLVLSFAATTPLADKKKPYFFRTVPSDNAVNPA 184
      173 -----SEKIGSEFLDMATTVAAGK-VELADCGKQIPST 208
QY      185 ILKLKHFRRRVGTLTDVQRFSEVRND-----LGVLTGGEIEISDTESFSNDPCT 237
Db      209 WCADSKGNPSTDTQVVLHGGGLPLGIEETGSKYKGTGLSMNGELFCGLIAGSSFGKNVR 268
      238 SVKRLKGN---DVRIL-----LGQFPDQ-----NMAAKYFCCAFEESSMFGSKYQ 277
QY      269 -----LWGQSHKADNGQC-----FVALDQCFAPGAFAPRLQOFLDETR 307
Db      278 WIIPGWYEPAMWEQYHVEANSRCLRSLLAMEGIVDFE----- 319
QY      308 NIKPISEKRPVLPG-DPERMNTYESQKAGL 338
Db      320 ---PLSSKQIKTISGKTPOQYEREYNSKRSGV 348

```

RESULT 11

```

US-09-211-755B-4
; Sequence 4, Application US/09211755B
; Patent No. US20020045742A1
; GENERAL INFORMATION:
; APPLICANT: Kenneth A. Jones, Thomas M. Laz, Beth Borowsky
; TITLE OF INVENTION: DNA Encoding a GABABR2 Polypeptide And Uses Thereof
; FILE REFERENCE: 1795/54002-D
; CURRENT APPLICATION NUMBER: US/09/211,755B
; CURRENT FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 09/186,664
; PRIOR FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-211-755B-4

```

Query Match 5.0%; Score 98; DB 9; Length 940;

Best Local Similarity 19.5%; Pred. No. 0.77;

Matches 53; Conservative 35; Mismatches 80; Indels 104; Gaps 12;

```

QY      123 NNY-----GACGCHYTKRIANA-----GWGMSFTNTSPLMPCRSSEIGLTNPSCVN-- 172
      125 NMLAVGVCPSVTSLIAESLOGMNLVLSFAATTPLADKKKPYFFRTVPSDNAVNPA 184
Db      173 -----SEKIGSEFLDMATTVAAGK-VELADCGKQIPST 208
      185 ILKLKHFRRRVGTLTDVQRFSEVRND-----LGVLTGGEIEISDTESFSNDPCT 237
QY      209 WCADSKGNPSTDTQVVLHGGGLPLGIEETGSKYKGTGLSMNGELFCGLIAGSSFGKNVR 268
      238 SVKRLKGN---DVRIL-----LGQFPDQ-----NMAAKYFCCAFEESSMFGSKYQ 277
Db      269 -----LWGQSHKADNGQC-----FVALDQCFAPGAFAPRLQOFLDETR 307
      278 WIIPGWYEPAMWEQYHVEANSRCLRSLLAMEGIVDFE----- 319
QY      308 NIKPISEKRPVLPG-DPERMNTYESQKAGL 338
      320 ---PLSSKQIKTISGKTPOQYEREYNSKRSGV 348

```

RESULT 12

```

US-09-779-334A-7
; Sequence 7, Application US/09779334A
; Publication No. US20030021774A1
; GENERAL INFORMATION:
; APPLICANT: Stjoeholm, Carsten
; APPLICANT: Oestergaard, Peter Rahbek
; APPLICANT: Kluentner, Anne-Marie
; TITLE OF INVENTION: Use of Acid-Stable Subtilisin Proteases in Animal Feed

```

```

; FILE REFERENCE: NOV1 100
; CURRENT APPLICATION NUMBER: US/09/779,334A
; CURRENT FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus sp. THS-1001
US-09-779-334A-7

```

Query Match 4.9%; Score 94.5; DB 11; Length 269;

Best Local Similarity 27.1%; Pred. No. 0.25;

Matches 51; Conservative 22; Mismatches 78; Indels 37; Gaps 10;

```

QY      68 GYKNGCVPKYLKQKGSTAVDGENL-----LGAVVGNFCTDLAIKLAKFQVAMVYTKN 121
      95 GANSGSVSSIAQ--GLQWTAQNNIHVANI.SLGSFVGSQTLAVQAATNAGVLVVAATG 152
Db      122 SNHYGACQHYTKRIANAGVMSFTNTSPLMPCRSSEIGLTNPISCCVNSEKT--GDS 179
      153 NNGSGTVS-YPARYANALAVGATDQNNRASF-----SQTGTQNTIVAPGVGIQSTPGNR 207
QY      180 FL-----LDMATTVAAGKVELADCGKQIPSTWGDADSKGNPS-TDTQVVLH-GGGILPL 233
      208 YASLSTGSMATPHVA-GVAALV-----KKNKNSMSNTQIRQHLTSTATSL 251
Db      234 GGIETGS 241
      252 GNSNOFGS 259

```

RESULT 13

```

US-09-813-408-7
; Sequence 7, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Maris, Barry
; APPLICANT: Delagrave, Simon
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-813-408-7

```

Query Match 4.9%; Score 94.5; DB 11; Length 382;

Best Local Similarity 27.1%; Pred. No. 0.44;

Matches 51; Conservative 22; Mismatches 78; Indels 37; Gaps 10;

```

QY      68 GYKNGCVPKYLKQKGSTAVDGENL-----LGAVVGNFCTDLAIKLAKFQVAMVYTKN 121
      208 GANSGSVSSIAQ--GLQWTAQNNIHVANI.SLGSFVGSQTLAVQAATNAGVLVVAATG 265
Db      122 SNHYGACQHYTKRIANAGVMSFTNTSPLMPCRSSEIGLTNPISCCVNSEKT--GDS 179
      266 NNGSGTVS-YPARYANALAVGATDQNNRASF-----SQTGTQNTIVAPGVGIQSTPGNR 320
QY      180 FL-----LDMATTVAAGKVELADCGKQIPSTWGDADSKGNPS-TDTQVVLH-GGGILPL 233
      321 YASLSTGSMATPHVA-GVAALV-----KKNKNSMSNTQIRQHLTSTATSL 364
Db      234 GGIETGS 241
      365 GNSNOFGS 372

```

Query Match	4.88;	Score 92.5;	DB 10;	Length 1017;
Best Local Similarity	18.6%;	Pred. No. 3.3;		
Matches 93; Conservative	66;	Mismatches 145;	Indels 195;	Gaps 25

RESULT 15
US-09-818-879-2

Sequence 2 Application US/09818879
Patent No. US20010023289A1
GENERAL INFORMATION:
APPLICANT: Jones, Kenneth
APPLICANT: Laz, Thomas
APPLICANT: Borowsky, Beth
TITLE OF INVENTION: DNA encoding a GABAR2 polypeptide and uses thereof
FILE REFERENCE: 1795/54002DA

Query Match	4.7%;	Score	92;	DB	9;	Length	898;
Best Local Similarity	18.8%;	Pred.	No. 3;				
Matches	51;	Conservative	37;	Mismatches	80;	Indels	104;
						Gaps	12

```

QY      123  NH-----GACQHTKKIAA-----GMVGSFNTSPLEPCRSSBIGTPTLSCVNA-- 172
D      83  NHTLVGGVCPSTSTIAESLQGMNLVQLSFAATTPVLADKKKPYEFRTYPSDMANPA 142
QY      173  -----SEKTDGDFLLDMATTVALGK-VELADCRGKTQIPST 208
D      143  IELKLHYOMKRVGLTQDVQRFSEVRND-----LTGVLYGDEIETSDTESNSDPC 199
QY      209  WGADSKGNPDTPQVAVLVHGGGLLPGLGIEETGSYKGTLSMAGELFCGILTAGSFGKNR 266
D      196  SVAKKLGN---DVRII-----LGQFDQ-----MNAKVFCAIEENMYGSKYQ 235
QY      269  -----LMGSHKADNGOC-----EVALDCEFAFGAPRLQGLFIDETR 307
D      236  WIIPGVYEESWMEQVHTEANSRCKRLNLAMEGIGVDPE----- 277
QY      308  NUKPISEKPVLYPG-DPERMNTETYSOKAGL 338
D      278  ---PLSSKQIKTISGKTPOQYERETNNKRSYV 306

```

Search completed: July 24, 2003, 21:10:55
Job time : 73 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:23:00 ; Search time 45 Seconds
(without alignments)
782.172 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 366

Sequence: 1 MNSKDAPEVVSPPKDAREF.....ALEELATKCVQMFSTKRLK 366

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	2.7	400	2	T18570
2	9	2.5	317	2	S01412
3	9	2.5	355	2	T20825
4	8	2.2	104	2	AH0040
5	8	2.2	205	2	G84204
6	8	2.2	299	2	AC0507
7	8	2.2	369	2	F96969
8	8	2.2	372	2	T20396
9	8	2.2	461	2	S21223
10	8	2.2	870	2	T01310
11	7	1.9	104	1	A28215
12	7	1.9	116	2	H90512
13	7	1.9	116	2	AH3150
14	7	1.9	133	2	G45853
15	7	1.9	146	2	S69694
16	7	1.9	174	2	AC1587
17	7	1.9	174	2	AG1652
18	7	1.9	197	2	B64410
19	7	1.9	201	2	H90347
20	7	1.9	208	2	E64357
21	7	1.9	210	2	D87467
22	7	1.9	221	2	AC1131
23	7	1.9	221	2	AC1491
24	7	1.9	227	2	H82005
25	7	1.9	229	2	S73777
26	7	1.9	230	2	D89821
27	7	1.9	231	2	T06258
28	7	1.9	231	2	T06801
29	7	1.9	231	2	T04072

30	7	1.9	231	2	T04075
31	7	1.9	231	2	T04312
32	7	1.9	232	2	C48684
33	7	1.9	233	2	A48684
34	7	1.9	233	2	B48684
35	7	1.9	234	2	A12709
36	7	1.9	234	2	B97492
37	7	1.9	235	2	S03839
38	7	1.9	241	2	A75065
39	7	1.9	243	2	G72482
40	7	1.9	244	2	C89884
41	7	1.9	244	2	AG2243
42	7	1.9	246	2	B90237
43	7	1.9	247	2	AG1672
44	7	1.9	253	2	T46830
45	7	1.9	253	2	AF0916

ALIGNMENTS

```

RESULT 1
T18570
probable malate dehydrogenase (EC 1.1.1.37) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18570; T20833
R:Almscough, R.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z18979
A:Accession: T18570
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-400 <MIL>
A:Cross-references: EMBL:AL033535; PIDN:CAA22133.1; CESP:VF13D12L.3
A:Experimental source: clone VF13D12L
R:Coles, L.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19330
A:Accession: T20833
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-400 <MIL>
A:Cross-references: EMBL:Z49127; PIDN:CAA88951.1; GSPDB:GN00020; CESP:VF13D12L.3
A:Experimental source: clone F13D12
C:Genetics:
A:Gene: CESP:VF13D12L.3
A:Map position: 2
A:Introns: 53/3; 96/1; 179/3; 230/3; 276/1; 315/3; 347/3; 384/3
C:Superfamily: malate dehydrogenase yIbc
C:Keywords: oxidoreductase

Query Match
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 GHYSHGLNRL 58
Db 86 GHYSHGLNRL 95

RESULT 2
S01412
hypothetical protein, 35K - red clover necrotic mosaic virus
C:Species: red clover necrotic mosaic virus
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 10-Dec-1999
C:Accession: S01412
R:Lommel, S.A.; Weston-Pina, M.; Xiong, Z.; Lomonosoff, G.P.
Nucleic Acids Res. 16, 8587-8602, 1988
A:Title: The nucleotide sequence and gene organization of red clover necrotic mosaic
A:Reference number: S01412; MUID:88335609; PMID:3047682
A:Accession: S01412
A:Molecule type: mRNA

```

A;Residues: 1-317 <LOM>
 A;Cross-references: EMBL:X08021; NID:g61465; PIDN:CAA30822.1; PID:g61466
 C;Genetics:
 A;Map position: RNA-2
 C;Superfamily: carnation ringspot virus cell movement protein

Query Match 2.5%; Score 9; DB 2; Length 317;
 Best Local Similarity 100.0%; Pred. No. 0.56;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 KYLKOKGT 84
 DB 245 KYLKOKGT 253

RESULT 3
 T20825
 hypothetical protein F13D12.1 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T20825
 R;Coles, L.

submitted to the EMBL Data Library, April 1995
 A;Reference number: Z19330
 A;Accession: T20825
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-355 <WTL>
 A;Cross-references: EMBL:Z49127; PIDN:CAA8943.1; GSPDB:GN00020; CESP:F13D12.1
 A;Experimental source: clone F13D12
 C;Genetics:
 A;Gene: CESP:F13D12.1
 A;Map position: 2
 A;Introns: 21/3; 67/1; 106/3; 138/3; 175/3; 189/3; 227/1; 258/3; 293/1

Query Match 2.5%; Score 9; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 0.62;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 227 GGGLPLPG 235
 DB 54 GGGLPLPG 62

RESULT 4
 AH0040
 conserved hypothetical protein YPO0326 [imported] - *Yersinia pestis* (strain CO92)
 C;Species: *Yersinia pestis*
 C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C;Accession: AH0040
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
 A;Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A;Reference number: AB0001; MOID:21470413; PMID:11586360
 A;Accession: AH0040
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-104 <KUR>
 A;Cross-references: GB:AL590842; PIDN:CAC89187.1; PID:g15978426; GSPDB:GN00175
 C;Genetics:
 A;Gene: YPO0326

Query Match 2.2%; Score 8; DB 2; Length 104;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 302 FLDETRNL 309
 DB 44 FLDETRNL 51

RESULT 5
 G84204
 hypothetical protein Vng0463c [imported] - *Halobacterium* sp. NRC-1
 C;Species: *Halobacterium* sp. NRC-1
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C;Accession: G84204
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
 A;Title: Genome sequence of *Halobacterium* species NRC-1.
 A;Reference number: AB4160; MOID:20504483; PMID:11016950
 A;Accession: G84204
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-205 <STO>
 A;Cross-references: GB:AE004437; NID:g10580069; PIDN:AAG19003.1; GSPDB:GN00138
 C;Genetics:
 A;Gene: VNG0463C

Query Match 2.2%; Score 8; DB 2; Length 205;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 VLHGGGL 231
 DB 158 VLHGGGL 165

RESULT 6
 AC0507
 transcription activator protein Nhar STY0048 [imported] - *Salmonella enterica* subsp.
 C;Species: *Salmonella enterica* subsp. *enterica* serovar Typh
 A;Note: this species has also been called *Salmonella typhi*
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AC0507
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr, S.; Moule, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se
 A;Reference number: AB0502; MOID:21534947; PMID:11677608
 A;Accession: AC0507
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-299 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD01195.1; PID:g16501324; GSPDB:GN00176
 C;Genetics:
 A;Gene: STY0048
 C;Superfamily: regulatory protein nhar

Query Match 2.2%; Score 8; DB 2; Length 299;
 Best Local Similarity 100.0%; Pred. No. 5.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 343 GOIKALEE 350
 DB 39 GOIKALEE 46

RESULT 7
 F96969
 malate dehydrogenase [imported] - *Clostridium acetobutylicum*
 C;Species: *Clostridium acetobutylicum*
 C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C;Accession: F96969
 R;Molling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L., Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A;Reference number: A96900; MOID:21359325; PMID:21359325

A:Accession: F66969
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-369 <KRP>
 A:Cross-references: GB:AF001437; PIDN:NAK78545.1; PID:g15023433; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0566
 C:Superfamily: malate dehydrogenase y1bc

Query Match 2.2%; Score 8; DB 2; Length 369;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 228 GGLPLPG 235
 |||||
 Db 232 GGLPLPG 239

RESULT 8
 T20396
 hypothetical protein F36A2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T20396; T21836
 R:Lennard, N.
 Submitted to the EMBL Data Library, June 1997
 A:Reference number: 219266
 A:Accession: T20396
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-372 <WIL>
 A:Cross-references: EMBL:Z96047; PIDN:CAB09417.1; GSPDB:GN00019; CESP:F36A2.3
 A:Experimental source: clone D13
 R:Lennard, N.
 Submitted to the EMBL Data Library, October 1996
 A:Reference number: 219476
 A:Accession: T21836
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-372 <WIL>
 A:Cross-references: EMBL:Z81077; PIDN:CAB03073.1; GSPDB:GN00019; CESP:F36A2.3
 A:Experimental source: clone F36A2
 C:Genetics:
 A:Gene: CESP:F36A2.3
 A:Map position: 1
 A:Introns: 21/3; 53/2; 79/1; 114/3; 147/3; 198/3; 249/1; 284/3; 353/3
 C:Superfamily: malate dehydrogenase y1bc

Query Match 2.2%; Score 8; DB 2; Length 372;
 Best Local Similarity 100.0%; Pred. No. 7.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 GQCFVAD 287
 |||||
 Db 285 GQCFVAD 292

RESULT 9
 S21223
 triacylglycerol lipase (EC 3.1.1.3) precursor, pancreatic - horse (fragment)
 C:Species: Equus caballus (domestic horse)
 C:Date: 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 18-Jun-1999
 C:Accession: S21223; S21251
 R:Kerfelec, B.; Foglitzo, E.; Bonicel, J.; Bougis, P.E.; Chapus, C.
 Eur. J. Biochem. 206, 279-287, 1992
 A:Title: Sequence of horse pancreatic lipase as determined by protein and cDNA sequencing
 A:Reference number: S21223; MUID:92267022; PMID:1587279
 A:Accession: S21223
 A:Molecule type: mRNA
 A:Residues: 1-461 <KER>
 A:Cross-references: EMBL:X66218; NID:g1063; PIDN:CAA46961.1; PID:g1064
 A:Accession: S21251

A:Molecule type: protein
 A:Residues: 13-78;80;97-218;220-228;230-277;322-349;384-438;440-448;450-461 <KEW>
 C:Superfamily: triacylglycerol lipase
 C:Keywords: carboxylic ester hydrolase
 F:1-12/Domain: signal sequence (fragment) #status predicted <SIG>
 F:13-461/Product: triacylglycerol lipase #status experimental <MAT>

Query Match 2.2%; Score 8; DB 2; Length 461;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 LGAVVGN 99
 |||||
 Db 6 LGAVVGN 13

RESULT 10
 T01310
 hypothetical protein T14P8.20 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 14-May-1999
 C:Accession: T01310
 R:Kallick, J.; Elliott, G.; Cloud, J.
 Submitted to the EMBL Data Library, May 1998
 A:Description: The sequence of A. thaliana T14P8.
 A:Reference number: Z14290
 A:Accession: T01310
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-870 <KAL>
 A:Cross-references: EMBL:AF069298; NID:g3193282; PID:g3193300
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4
 A:Introns: 89/3; 133/1; 296/3; 369/3; 474/3; 577/3; 604/3; 630/3; 757/3; 813/3
 A:Note: T14P8.20

Query Match 2.2%; Score 8; DB 2; Length 870;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 EKPVLPG 322
 |||||
 Db 698 EKPVLPG 705

RESULT 11
 A28215
 thioredoxin - Rhodospirillum rubrum
 C:Species: Rhodospirillum rubrum
 C:Date: 28-Aug-1989 #sequence_revision 26-May-1994 #text_change 11-Apr-1995
 C:Accession: A28215
 R:Johnson, T.C.; Yee, B.C.; Carlson, D.E.; Buchanan, B.B.; Johnson, R.S.; Mathews, W.
 J. Bacteriol. 170, 2406-2408, 1988
 A:Title: Thioredoxin from Rhodospirillum rubrum: primary structure and relation to th
 A:Reference number: A28215; MUID:88198045; PMID:3129411
 A:Accession: A28215
 A:Molecule type: Protein
 A:Residues: 1-104 <JOH>
 C:Superfamily: thioredoxin; thioredoxin homology
 C:Keywords: redox-active disulfide
 F:7-90/Domain: thioredoxin homology <TPH>
 F:29-32/Disulfide bonds: redox active #status predicted

Query Match 1.9%; Score 7; DB 1; Length 104;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 347 ALEELAT 353
 |||||
 Db 38 ALEELAT 44

RESULT 12

H90512

hit-like protein (cell cycle regulation) [Imported] - Mycoplasma pulmonis (strain UAB CT

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C:Accession: H90512

R:Chamaud, I.; Heilbrunn, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallissot, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: H90512

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-116 <KOR>

A:Cross-references: GB:AL445566; PID:g14089421; PIDN:CAC13181.1; GSPDB:GN00153

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MPRU_0080

A:Genetic code: SGC3

C:Superfamily: protein kinase C inhibitor; histidine triad homology

Query Match

Best Local Similarity 1.9%; Score 7; DB 2; Length 116;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 AIKLAKE 111

DB 68 AIKLAKE 74

RESULT 13

AH3150

hypothetical protein Atu4828 [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AH3150

R:Wood, D.W.; Seubald, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608530; PMID:11743193

A:Accession: AH3150

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-116 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA145622.1; PID:g17743344; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4828

A:Map position: linear chromosome

Query Match

Best Local Similarity 1.9%; Score 7; DB 2; Length 116;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 344 QIKALEE 350

DB 23 QIKALEE 29

RESULT 14

G45893

T-cell receptor alpha chain precursor V region (BTA15) - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000

C:Accession: G45893

R:Ishiguro, N.; Tanaka, A.; Shinagawa, M.

Immunogenetics 31, 57-60, 1990

A:Title: Sequence analysis of bovine T-cell receptor alpha chain.

A:Reference number: A45893; MUID:90129157; PMID:2137108

A:Accession: G45893

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-133 <ISH>

A:Cross-references: GB:D90016; NID:g217620; PIDN:BA14066.1; PID:g217621

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: T-cell receptor

Query Match

Best Local Similarity 1.9%; Score 7; DB 2; Length 133;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 PGFAPRL 299

DB 64 PGFAPRL 70

RESULT 15

S69694

hypothetical protein YDR397C - Yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 24-Sep-1999

C:Accession: S69694; S69680

R:Dietrich, F.S.

submitted to the EMBL Data Library, July 1995

A:Description: The sequence of S. cerevisiae cosmid 9481, 9509, 9926, 9461, and lamb

A:Reference number: S69655

A:Accession: S69694

A:Molecule type: DNA

A:Residues: 1-146 <DIE>

A:Cross-references: EMBL:U32274; MIPS:YDR397C

A:Accession: S69680

A:Molecule type: DNA

A:Residues: 'MCRPRMTL', 13-146 <DIW>

A:Cross-references: EMBL:U32274; NID:g927313; PIDN:AB64838.1; PID:g927329

C:Genetics:

A:Gene: SGD:NCB2

A:Cross-references: SGD:S0002805; MIPS:YDR397C

A:Map position: 4R

A:Insertions: 13/1

C:Superfamily: conserved hypothetical protein YDR397C

Query Match

Best Local Similarity 1.9%; Score 7; DB 2; Length 146;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 IKALEEL 351

DB 70 IKALEEL 76

Search completed: July 24, 2003, 21:28:44
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:16:20 ; Search time 23 Seconds

(without alignments)
748.338 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 366
Sequence: 1 MNVSKDAPEFVVSPPKDAREF.....ALEELATKCDVQMFYKRLK 366

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.5	176	1 CG38_HUMAN	Q9bW30 homo sapien
2	9	2.5	176	1 CG38_MOUSE	Q9cD30 mus musculu
3	9	2.5	317	1 V35K_RCNMV	P10838 red clover
4	8	2.2	461	1 L1P2_HORSE	P29183 equus cabal
5	7	1.9	104	1 TH10_RHORI	P10473 rhodospiril
6	7	1.9	197	1 Y882_METJA	O58292 methanococc
7	7	1.9	208	1 RS3_METJA	P54034 methanococc
8	7	1.9	229	1 Y268_MYCPN	P75396 mycoplasma
9	7	1.9	231	1 SODM_ORYSA	Q43008 oryza sativ
10	7	1.9	232	1 SODN_MAIZE	P41979 zea mays (m
11	7	1.9	233	1 SODP_MAIZE	P41980 zea mays (m
12	7	1.9	233	1 SODP_MAIZE	P09233 zea mays (m
13	7	1.9	235	1 SODP_MAIZE	O33808 salmonella
14	7	1.9	252	1 UDP_SALT1	O08444 klebsiella
15	7	1.9	253	1 UDP_KLEAE	P02915 salmonella
16	7	1.9	258	1 H1SP_SALT1	O91850 archaeglob
17	7	1.9	264	1 TH1K_STACA	P52671 klebsiella
18	7	1.9	279	1 UDP_KLEPN	O9X309 coxiella bu
19	7	1.9	296	1 EFTS_COXBU	P52692 vibrio chol
20	7	1.9	296	1 NHAR_VIBCH	O27984 archaeoglob
21	7	1.9	298	1 DHY2_ARCFU	O88693 shewanella
22	7	1.9	304	1 RDGC_SHEON	O28550 archaeglob
23	7	1.9	305	1 YH24_ARCFU	P40870 bacillus su
24	7	1.9	323	1 SP55_BACSU	P54751 m cmp-n-ace
25	7	1.9	337	1 S14A_MOUSE	Q11201 h cmp-n-ace
26	7	1.9	340	1 S14A_MOUSE	Q02745 s cmp-n-ace
27	7	1.9	360	1 S14A_PIG	O59028 pyrococcus
28	7	1.9	363	1 MDR_PYROH	O9V065 pyrococcus
29	7	1.9	362	1 MDR_PYROH	P24217 e peps syste
30	7	1.9	376	1 PTRP_ECOLI	P48195 lactobacilli
31	7	1.9	427	1 PYRC_LACTE	O9V168 pyrococcus
32	7	1.9	443	1 SR54_PYRAB	O8u070 pyrococcus
33	7	1.9	443	1 SR54_PYRAB	

ALIGNMENTS

RESULT 1	CG38_HUMAN	STANDARD	PRT	176 AA.
AC	O9BM30; Q9Y326; Q9Y6H0;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	CGI-38 protein.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20272150; PubMed=10810093;			
RA	Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;			
RT	"Identification of novel human genes evolutionarily conserved in			
RT	Cenorchaditis elegans by comparative proteomics."			
RL	Genome Res. 10:703-713(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pituitary;			
RX	MEDLINE=20402571; PubMed=10931946;			
RA	Hu R.-M., Han Z.-G., Song Y.-D., Peng Y.-D., Huang Q.-H., Ren S.-X.,			
RA	Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H.,			
RA	Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,			
RA	Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,			
RA	Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.;			
RT	"Gene expression profiling in the human hypothalamus-pituitary-adrenal			
RT	axis and full-length cDNA cloning."			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Eye;			
RX	MEDLINE=22386257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stampleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Milllany S.J.,			
RA	Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kelleman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Butterfield Y.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.C., Krzywinski M.I., Skalska U., Smallos D.E.,			
RA	Scherer A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
CC	-1- SIMILARITY: BELONGS TO THE P25 FAMILY.			

CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 CC in position 52, 76, 83 and 88.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF132972; AAD27747.1; -
 DR EMBL: AF078846; AAD44478.1; ALT_FRAME.
 DR EMBL: BC000691; AAH00691.1; -
 FT CONFLICT 7 7 M -> I (IN REF. 2).
 FT CONFLICT 10 11 LE -> PK (IN REF. 2).
 FT CONFLICT 88 88 K -> Q (IN REF. 1).
 SO SEQUENCE 176 AA; 18985 MW; 20EAl46E3FBIADf CRC64;
 Query Match 2.5%; Score 9; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 346 KALEELATK 354
 Db 77 KALEELATK 85
 RESULT 2
 CG38_MOUSE STANDARD; PRT; 176 AA.
 AC Q9CRB6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE CGT-38 protein homolog.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo, and Tongue;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aikawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Flieschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
 RA Lyons P., Matchanoni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nardone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wliding L.,
 RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontuski S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=22386257; Pubmed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Donaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Brown S.S., Loguélano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schenker A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SIMILARITY: BELONGS TO THE P25 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AK009771; BAB26493.1; -
 DR EMBL: AK012437; BAB28237.1; -
 DR EMBL: BC010788; AAH10788.1; -
 DR MGD: MGI:1915221; 2700055K07RLK.
 SO SEQUENCE 176 AA; 18965 MW; 05CD1F57E263B156 CRC64;
 Query Match 2.5%; Score 9; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 346 KALEELATK 354
 Db 77 KALEELATK 85
 RESULT 3
 V35K_RCNMV STANDARD; PRT; 317 AA.
 AC P10838;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 35 kDa protein in RNA2.
 OS Red clover necrotic mosaic virus (RCNMV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
 CC Dianthovirus.
 OX NCBI_TaxID=12267;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Australian;
 RX MEDLINE=88335609; Pubmed=3047682;
 RA Lommel S.A., Weston Fina M., Xiong Z., Lomonosoff G.P.;
 RT "The nucleotide sequence and gene organization of red clover necrotic
 RT mosaic virus RNA-2";
 RL Nucleic Acids Res. 16:8587-8602(1988).
 CC -1- FUNCTION: RNA-2 MAY ENCODE A CELL-TO-CELL MOVEMENT FUNCTION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X08021; CAA30822.1; -
 DR PIR: S01412; S01412.
 DR InterPro: IPR000603; 3A_mov.

FT TURN 452 453
 FT STRAND 456 461
 SQ SEQUENCE 461 AA; 50921 MW; 382F33F3CE446738 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 8; DB 1; Length 461;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 92 LGAAGVGN 99
 Db 6 LGAAGVGN 13
 RESULT 5
 THIO_RHORU STANDARD; PRT; 104 AA.
 ID THIO_RHORU
 AC P10473;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thioresoxin (TRX).
 GN TRXA.
 OS Rhodospirillum rubrum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Rhodospirillaceae; Rhodospirillum.
 OX NCBI_TaxID=1085;
 RN [1]
 RP MEDLINE=88198045; PubMed=3129411;
 RA Johnson T.C., Yee B.C., Carlson D.E., Buchanan B.B., Johnson R.S.,
 RA Mathews W.R., Blemann K.,
 RT "Thioresoxin from Rhodospirillum rubrum: primary structure and
 RL relation to thioresoxins from other photosynthetic bacteria.";
 CC J. Bacteriol. 170:2406-2408(1988).
 CC -1- FUNCTION: Participates in various redox reactions through the
 CC reversible oxidation of its active center dithiol to a disulfide
 CC and catalyzes dithiol-disulfide exchange reactions.
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -1- CAUTION: X'S IN THE SEQUENCE ARE EITHER ILE OR LEU.
 CC PIR: A28215; A28215.
 DR HSSP: P00274; 177P.
 DR InterPro: IPR006662; Thioresoxin.
 DR InterPro: IPR005746; Thioresoxin_dom2.
 DR InterPro: IPR006663; Thioresoxin.
 DR Pfam: PF00085; Thioresoxin.
 DR TIGRFAMs: TIGR01068; Thioresoxin; 1.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KM Redox-active center; Electron transport.
 FT DISULFID 29 REDOX-ACTIVE (BY SIMILARITY).
 SQ SEQUENCE 104 AA; 11314 MW; 4A31E015FD71AE03 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 104;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 347 ALLELAT 353
 Db 38 ALLELAT 44
 RESULT 6
 Y882_METJA STANDARD; PRT; 197 AA.
 ID Y882_METJA
 AC Q58292;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein M00882.
 GN M00882.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=9637999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RL jannaschii.";
 RL Science 273:1058-1073(1996).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF A-197.
 RA Hung L.W., Huang L., Kim R., Kim S.H.;
 RT "Crystal structure and functional analysis of a hypothetical protein,
 RL M00882, from Methanococcus jannaschii.";
 CC Submitted (JAN-2000) to the PDB data bank.
 CC -1- SIMILARITY: BELONGS TO THE HEMK FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U67532; AAB98886.1; -
 DR PIR: B64410; B64410.
 DR PDB: 1DUS; 19-JUL-00.
 DR TIGR: M00882; -
 DR InterPro: IPR002052; N6_Mtase.
 DR InterPro: IPR000051; SAM_Dbind.
 DR Pfam: PF05175; MTS; 1.
 KM Transferrase; Methyltransferase; Complete proteome; 3D-structure.
 FT STRAND 14 21
 FT TURN 22 23
 FT STRAND 24 31
 FT TURN 32 33
 FT TURN 35 38
 FT TURN 42 50
 FT TURN 55 56
 FT STRAND 58 62
 FT TURN 63 63
 FT TURN 65 66
 FT TURN 68 73
 FT HELIX 74 76
 FT STRAND 79 84
 FT TURN 87 99
 FT HELIX 100 101
 FT TURN 103 104
 FT STRAND 107 111
 FT TURN 114 117
 FT TURN 119 120
 FT STRAND 123 128
 FT TURN 133 134
 FT HELIX 136 149
 FT STRAND 150 161
 FT TURN 163 163
 FT HELIX 164 177
 FT STRAND 181 186
 FT TURN 187 188
 FT STRAND 189 195
 SQ SEQUENCE 197 AA; 22244 MW; F4E108E500ABEB28 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 197;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 AIRLAKE 111
 |||||
 DB 89 AIRLAKE 95

RESULT 7
 RS3_METUA STANDARD: PRT: 208 AA.
 ID RS3_METUA
 AC P54034:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S3P.
 GN RPS3P OR M10461.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=9633799; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kertavagae A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geopghagen N.S.M., Meldman J.F., Fuhmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hrust M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL -1- FUNCTION: Binds the lower part of the 30S subunit head (By
 similarity).
 CC -1- SUBUNIT: Part of the 30S ribosomal subunit.
 CC -1- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
 CC -1- SIMILARITY: Contains 1 KH type-2 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL: U67497; AAB98450.1; -;
 DR PIR: E64357; E64357.
 DR TIGR: M10461; -;
 DR HAMAP: MF_01309; -; 1.
 DR InterPro: IPR004087; KH_dom.
 DR InterPro: IPR004044; KH_type_2.
 DR InterPro: IPR001351; Ribosomal_S3.
 DR InterPro: IPR005703; S3_euk_arch.
 DR Pfam: PF00013; KH; 1.
 DR Pfam: PF00189; Ribosomal_S3_C; 1.
 DR SMART: SM00322; KH; 1.
 DR TIGRFAMS: TIGR01008; ipsc_E_A; 1.
 DR PROSITE: PS50823; KH_type_2; 1.
 DR PROSITE: PS00548; RIBOSOMAL_S3; 1.
 KM Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
 FT DOMAIN 16
 FT SEQUENCE 208 AA: 23325 MW; C78DBE96A5A22A6A CRC64;

Query Match 1.9%; Score 7; DB 1; Length 208;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 LAKEGV 114
 |||||
 DB 68 LAKEGV 74

RESULT 8
 Y268_MYCPN STANDARD: PRT: 229 AA.
 ID Y268_MYCPN
 AC P75396;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MG268 homolog (r11_orf229).
 GN MEN386 OR M451.
 OS Mycoplasma pneumoniae.
 OC Bacteria: Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948653;
 RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
 RA Hermann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 pneumoniae.";
 RT Nucleic Acids Res. 24:4420-4449(1996).
 RL -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL: AE000044; AAB96099.1; -;
 DR PIR: S73777; S73777.
 KM Hypothetical protein; ATP-binding; Complete proteome.
 FT NP_BIND 22
 FT SEQUENCE 229 AA: 27020 MW; 97A17C27F9D25425 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 229;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 QUADLL 43
 |||||
 DB 53 QUADLL 59

RESULT 9
 SODM_ORYSA STANDARD: PRT: 231 AA.
 ID SODM_ORYSA
 AC Q43008;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
 GN SODA OR RMSOD1.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Etolated shoot;
 RA Pan S.M., Chen J.C.;
 RT "Rice manganese superoxide dismutase are encoded by multigene
 family.";
 RT Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 RL -1- FUNCTION: Destroys radicals which are normally produced within the
 cells and which are toxic to biological systems.
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -1- COFACTOR: Manganese (By similarity).
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.

```

CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L34038; AAA57130.1; -.
CC PIR; T04072; T04072.
CC HSSP; P04179; IAP6.
CC ANU-2DPAGE; Q43008; -.
CC Gramene; Q43008; -.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sode; 1.
CC DR Pfam; PF02777; sode; 1.
CC DR ProDom; PD000475; SODismutase; 1.
CC DR PROSITE; PS00088; SOD_MN; 1.
CC KM Oxidoreductase; Manganese; Mitochondrion; Transit peptide.
CC TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN 28 231 SUPEROXIDE DISMUTASE [MN].
CC FT METAL 59 59 MANGANESE (BY SIMILARITY).
CC FT METAL 103 103 MANGANESE (BY SIMILARITY).
CC FT METAL 192 192 MANGANESE (BY SIMILARITY).
CC FT METAL 196 196 MANGANESE (BY SIMILARITY).
CC SQ SEQUENCE 231 AA; 24947 MW; 0FCFD08FC27825C0 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 NLRPISE 314
Db 109 NLRPISE 115

RESULT 10
SODN_MAIZE STANDARD; PRT; 232 AA.
AC P41976;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Mn] 3.2, mitochondrial precursor (EC 1.15.1.1).
GN SODA.1 OR SOD3.2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94022365; Pubmed=8415698;
RA Zhu D., Scandalios J.G.;
RT "Maize mitochondrial manganese superoxide dismutases are encoded by a
RT differentially expressed multigene family."
RT Proc. Natl. Acad. Sci. U.S.A. 90:9310-9314(1993).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Manganese.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L19461; AAA72020.2; -.
CC PIR; C48684; C48684.
CC HSSP; P04179; IAP6.
CC DR MalzEDB; 47587; -.
CC DR InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sode; 1.
CC DR Pfam; PF02777; sode; 1.
CC DR ProDom; PD000475; SODismutase; 1.
CC DR PROSITE; PS00088; SOD_MN; 1.
CC KM Oxidoreductase; Manganese; Mitochondrion; Transit peptide;
CC Multigene family.
CC TRANSIT 1 29 MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN 30 232 SUPEROXIDE DISMUTASE [MN] 3.2.
CC FT METAL 57 57 MANGANESE (BY SIMILARITY).
CC FT METAL 105 105 MANGANESE (BY SIMILARITY).
CC FT METAL 193 193 MANGANESE (BY SIMILARITY).
CC FT METAL 197 197 MANGANESE (BY SIMILARITY).
CC SQ SEQUENCE 232 AA; 25357 MW; 6A0BB981281DB468 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 232;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 NLRPISE 314
Db 111 NLRPISE 117

RESULT 11
SODN_MAIZE STANDARD; PRT; 233 AA.
AC P41979;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Mn] 3.3, mitochondrial precursor (EC 1.15.1.1).
GN SODA.2 OR SOD3.3.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94022365; Pubmed=8415698;
RA Zhu D., Scandalios J.G.;
RT "Maize mitochondrial manganese superoxide dismutases are encoded by a
RT differentially expressed multigene family."
RT Proc. Natl. Acad. Sci. U.S.A. 90:9310-9314(1993).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Manganese.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE EMBRYO LATE IN
CC EMBRYOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR PIR; A48684; A48684.
 DR HSSP; P04179; 1ABM.
 DR Maizedb; 47587; -
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe_1.
 DR Pfam; PF02777; sodfe_C; 1.
 DR ProDom; PD000475; SODismutase; 1.
 DR PROSITE; PS00088; SOD_MN; 1.
 KW Oxidoreductase; Manganese; Mitochondrion; Transit peptide;
 MultiGene family.
 KW TRANSIT 1 29
 FT CHAIN 30 233 MITOCHONDRION (BY SIMILARITY).
 FT METAL 57 57 SUPEROXIDE DISMUTASE [MN] 3.3.
 FT METAL 105 105 MANGANESE (BY SIMILARITY).
 FT METAL 194 194 MANGANESE (BY SIMILARITY).
 FT METAL 198 194 MANGANESE (BY SIMILARITY).
 FT METAL 198 198 MANGANESE (BY SIMILARITY).
 SQ SEQUENCE 233 AA; 25448 MW; 061160B9B195A96 CRC64;

Query Match Best Local Similarity 1.9%; Score 7; DB 1; Length 233;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 NIKPISE 314
 Db 111 NIKPISE 117

RESULT 12
 SODM_MAIZE STANDARD; PRT; 233 AA.
 ID SODM_MAIZE
 AC P41980;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Superoxide dismutase [Mn] 3.4, mitochondrial precursor (EC 1.15.1.1).
 GN SODA.3 OR SOD3.4.
 OS Zea mays (Maize).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OK NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94022365; PubMed=8415698;
 RA Zhu D., Scandalios J.G.,
 RT "Maize mitochondrial manganese superoxide dismutases are encoded by a
 RT differentially expressed multigene family."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:9310-9314(1993).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Manganese.
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; L19463; AA72022.2; -
 DR PIR; B48684; B48684.
 DR HSSP; P04179; 1ABP.
 DR Maizedb; 47587; -
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe_1.
 DR Pfam; PF02777; sodfe_C; 1.
 DR ProDom; PD000475; SODismutase; 1.

DR PROSITE; PS00088; SOD_MN; 1.
 KW Oxidoreductase; Manganese; Mitochondrion; Transit peptide;
 MultiGene family.
 KW TRANSIT 1 29
 FT CHAIN 30 233 MITOCHONDRION (BY SIMILARITY).
 FT METAL 57 57 SUPEROXIDE DISMUTASE [MN] 3.4.
 FT METAL 105 105 MANGANESE (BY SIMILARITY).
 FT METAL 194 105 MANGANESE (BY SIMILARITY).
 FT METAL 194 194 MANGANESE (BY SIMILARITY).
 FT METAL 198 198 MANGANESE (BY SIMILARITY).
 SQ SEQUENCE 233 AA; 25239 MW; 214626404EB742A3 CRC64;

Query Match Best Local Similarity 1.9%; Score 7; DB 1; Length 233;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 NIKPISE 314
 Db 111 NIKPISE 117

RESULT 13
 SODM_MAIZE STANDARD; PRT; 235 AA.
 ID SODM_MAIZE
 AC P09233;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Superoxide dismutase [Mn] 3.1, mitochondrial precursor (EC 1.15.1.1).
 GN SODA.4 OR SOD3.1 OR SOD3.
 OS Zea mays (Maize).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OK NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. W64A;
 RX MEDLINE=89051020; PubMed=2461225;
 RA Redinbaugh M.G., Wadsworth G.T., Scandalios J.G.,
 RT "Isolation and characterization of a cDNA for mitochondrial manganese
 RT superoxide dismutase (SOD-3) of maize and its relation to other
 RT manganese superoxide dismutases."
 RL Biochim. Biophys. Acta 951:61-70(1988).
 CC -!- FUNCTION: Destroys radicals which are normally produced within the
 CC cells and which are toxic to biological systems.
 CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
 CC -!- COFACTOR: Manganese.
 CC -!- SUBUNIT: Homotrimer.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; X12540; CAA31058.1; -
 DR HSSP; P04179; 1ABM.
 DR Maizedb; 47587; -
 DR InterPro; IPR001189; SODismutase.
 DR Pfam; PF00081; sodfe_1.
 DR Pfam; PF02777; sodfe_C; 1.
 DR ProDom; PD000475; SODismutase; 1.
 DR PROSITE; PS00088; SOD_MN; 1.
 KW Oxidoreductase; Manganese; Mitochondrion; Transit peptide;
 MultiGene family.
 KW TRANSIT 1 31
 FT CHAIN 32 235 MITOCHONDRION.
 FT METAL 59 59 SUPEROXIDE DISMUTASE [MN] 3.1.
 FT METAL 59 59 MANGANESE (BY SIMILARITY).

FT METAL 107 107 MANGANESE (BY SIMILARITY).
 FT METAL 196 196 MANGANESE (BY SIMILARITY).
 FT METAL 200 200 MANGANESE (BY SIMILARITY).
 SO SEQUENCE 235 AA; 25545 MW; AD51BAD0F44FDE56 CRC64;
 Query Match 1.98; Score 7; DB 1; Length 235;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 308 NLRPSE 314
 Db 113 NLRPSE 119
 RESULT 14
 UDP_SALTY STANDARD; PRT; 252 AA.
 AC 033808; 008432; 0916M8;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Uridine phosphorylase (EC 2.4.2.3) (UDRPhase).
 GN UDP OR STM3968 OR STM01.21 OR STY3591 OR T3329.
 OS Salmonella typhimurium, and
 OC Salmonella typhi.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 ON NCBI_TaxID=602, 601;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhimurium; STRAIN-LT2;
 RA Errais L.B., Ukhadolina L.S., Eremina S.Y., Evdokimova A.A.,
 RA Mironov A.S.;
 RT "Structure and expression of the gene encoding uridine phosphorylase
 RT (udp) in Salmonella typhimurium";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhimurium; STRAIN-LT2;
 RX MEDLINE=98326439; PubMed=9661793;
 RA Velko V.P., Chebotayev D.V., Ovcharova I.V., Gul'ko L.B.;
 RA "Protein engineering of uridine phosphorylase from Escherichia coli
 RT K-12. I. Cloning and expression of uridine phosphorylase genes from
 RT Klebsiella aerogenes and Salmonella typhimurium in E. coli.";
 RL Bioorg. Khim. 24:381-387(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhimurium; STRAIN-LT2 / SSGC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhi; STRAIN-CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahina M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 RN [5]

RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhi; STRAIN-TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: THE ENZYMES WHICH CATALYZE THE REVERSIBLE PHOSPHORYLOSIS
 CC OF PYRIMIDINE NUCLEOSIDES ARE INVOLVED IN THE DEGRADATION OF THESE
 CC COMPOUNDS AND IN THEIR UTILIZATION AS CARBON AND ENERGY SOURCES,
 CC OR IN THE RESCUE OF PYRIMIDINE BASES FOR NUCLEOTIDE SYNTHESIS (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Uridine + phosphate = uracil + alpha-D-ribose
 CC 1-phosphate.
 CC -1- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
 CC -1- SUBUNIT: Homohexamer (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PNP/UDP FAMILY 1 OF PHOSPHORYLASES.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y14282; CAAT4658.2; -
 DR EMBL: Y13360; CAAT73795.1; -
 DR EMBL: AF233324; AAF33424.1; -
 DR EMBL: AE008885; AAL22812.1; -
 DR EMBL: AL627278; CAD07924.1; -
 DR EMBL: AE016845; AAO70857.1; -
 DR STGene; SG10731; udp.
 DR InterPro: IPR000845; PNP_UDP.
 DR Pfam: PF01048; PNP_UDP_1; 1.
 DR PROSITE: PS01232; PNP_UDP_1; 1.
 DR TRANSFERASE; Glycosyltransferase; Complete proteome.
 FT INT_MET 0
 FT CONFLICT 0
 FT SEQUENCE 199 199 A -> R (IN REF. 2).
 SO SEQUENCE 252 AA; 27008 MW; 5A3B6AF961EBED2 CRC64;
 Query Match 1.98; Score 7; DB 1; Length 235;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 320 VPDPPE 326
 Db 23 VPDPPE 29
 RESULT 15
 UDP_KLEAE STANDARD; PRT; 253 AA.
 AC 008444;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Uridine phosphorylase (EC 2.4.2.3) (UDRPhase).
 GN UDP.
 OS Klebsiella aerogenes.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Klebsiella.
 ON NCBI_TaxID=28451;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. typhi; STRAIN-CT18;
 RX MEDLINE=98326439; PubMed=9661793;
 RA Velko V.P., Chebotayev D.V., Ovcharova I.V., Gul'ko L.B.;
 RA "Protein engineering of uridine phosphorylase from Escherichia coli
 RT K-12. I. Cloning and expression of uridine phosphorylase genes from
 RT Klebsiella aerogenes and Salmonella typhimurium in E. coli.";
 RL Bioorg. Khim. 24:381-387(1998).
 RN [5]

CC -1- FUNCTION: THE ENZYMES WHICH CATALYZE THE REVERSIBLE PHOSPHORYLOSIS
 CC OF PYRIMIDINE NUCLEOSIDES ARE INVOLVED IN THE DEGRADATION OF THESE
 CC COMPOUNDS AND IN THEIR UTILIZATION AS CARBON AND ENERGY SOURCES.
 CC OR IN THE RESCUE OF PYRIMIDINE BASES FOR NUCLEOTIDE SYNTHESIS (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Uridine + phosphate -> uracil + alpha-D-ribose
 CC 1-phosphate.
 CC -1- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
 CC -1- SUBUNIT: Homohexamer (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE PNP/UDP FAMILY 1 OF PHOSPHORYLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: Y13414; CAA73838.1; -.
 CC DR PIR: T46830; T46830.
 CC DR InterPro: IPR000845; PNP_UDP.
 CC DR Pfam: PF01048; PNP_UDP_1; 1.
 CC DR PROSITE: PS01232; PNP_UDP_1; 1.
 CC KM Transferase: Glycosyltransferase.
 CC SQ SEQUENCE 253 AA; 27034 MW; 617A62F00386A405 CRC64;

Query Match 1.9%; Score 7; DB 1; Length 253;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 320 VFGDPER 326
 |||||
 DB 24 VFGDPER 30

Search completed: July 24, 2003, 21:26:02
 Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 21:16:25 ; Search time 103 Seconds

(without alignments)
916.964 Million cell updates/sec

Title: US-10-060-848-3

Perfect score: 366
Sequence: 1 MNYSKDAPEFVSPKDAREF.....ALEELATKCDVQMSYKRLK 366

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	2.7	400	5	Q9XTB4
2	8	2.2	104	16	Q8ZJ05
3	8	2.2	105	16	Q8DIF1
4	8	2.2	205	17	Q9HS05
5	8	2.2	299	16	Q8XG3
6	8	2.2	361	5	Q8IP19
7	8	2.2	369	16	Q97LJ4
8	8	2.2	372	5	Q62128
9	8	2.2	381	5	Q8IPU0
10	8	2.2	416	2	Q8KUC9
11	8	2.2	446	5	Q961U7
12	8	2.2	446	5	Q9VP68
13	8	2.2	474	2	Q8GGM7
14	8	2.2	523	16	Q8XK17
15	8	2.2	651	10	Q8VX51
16	8	2.2	676	10	Q8VX50

17	8	2.2	676	10	Q8VX52	Q8VX52 solanum tub
18	8	2.2	870	10	Q81293	Q81293 arabidopsis
19	8	2.2	1162	16	Q8DBL9	Q8DBL9 vibrio vuln
20	7	1.9	32	2	Q05663	Q05663 lactobacilli
21	7	1.9	82	2	Q05354	Q05354 wolbachia s
22	7	1.9	82	2	Q05358	Q05358 wolbachia s
23	7	1.9	91	2	Q9EXM1	Q9EXM1 escherichia
24	7	1.9	105	16	Q98P93	Q98P93 rhizobium l
25	7	1.9	105	16	Q983Y0	Q983Y0 rhizobium l
26	7	1.9	112	16	Q8RHM7	Q8RHM7 fusobacteri
27	7	1.9	116	16	Q8U6H9	Q8U6H9 agrobacteri
28	7	1.9	116	16	Q98RKO	Q98RKO mycoplasma
29	7	1.9	121	16	Q9F3P5	Q9F3P5 streptomyce
30	7	1.9	123	2	Q9ZAN6	Q9ZAN6 comamonas s
31	7	1.9	123	17	Q8TZ43	Q8TZ43 methanopyru
32	7	1.9	128	17	Q8U467	Q8U467 pyrococcus
33	7	1.9	143	3	Q04176	Q04176 saccharomyc
34	7	1.9	146	3	Q92317	Q92317 saccharomyc
35	7	1.9	146	4	Q9HBT0	Q9HBT0 homo sapien
36	7	1.9	146	4	Q8NC13	Q8NC13 homo sapien
37	7	1.9	149	2	Q93QML	Q93QML erythra tra
38	7	1.9	150	10	Q9ARW3	Q9ARW3 oryza sativ
39	7	1.9	158	11	Q9CZY0	Q9CZY0 mus musculu
40	7	1.9	174	16	Q92CD6	Q92CD6 listeria in
41	7	1.9	174	16	Q92B08	Q92B08 listeria in
42	7	1.9	178	16	Q8EWS5	Q8EWS5 mycoplasma
43	7	1.9	180	11	Q8BJX5	Q8BJX5 mus musculu
44	7	1.9	184	2	Q937M6	Q937M6 photorhabdu
45	7	1.9	190	5	Q8MRRO	Q8MRRO drosophila

ALIGNMENTS

RESULT 1

Q9XTB4

PRELIMINARY: PRT: 400 AA.

AC Q9XTB4:

01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE VF13D12L.3 protein.

GN VF13D12L.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Coles L.;

RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Johnston L.,

RA Garner A., Kershaw J., Kilsten J., Laiter N., Latreille P.,

RA Jones M., Kershaw J., Kilsten J., Laiter N., Latreille P.,

RA Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showken R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vandin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RL Nature 368:32-38(1994).

RN [3]

RP SEQUENCE FROM N.A.

RA Ainscough R.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z49127; CAA88951.1; -

DR EMBL: AL033535; CAA88951.1; JOINED.

DR EMBL; AL033535; CAA22133.1; -
 DR EMBL; Z49127; CAA22133.1; JOINED.
 DR WormPep; VF13D12L.3; CE21229.
 DR InterPro; IPR003767; Idb_2.
 DR Pfam; PF02615; Idb_2; 1
 SO SEQUENCE 400 AA; 42827 MW; 75C3E19CDEE5BDD CRC64;

Query Match 2.7%; Score 10; DB 5; Length 400;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GHYSHGLNRL 58
 DB 86 GHYSHGLNRL 95

RESULT 2

ID 082J05 PRELIMINARY; PRT; 104 AA.
 AC 082J05;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein YP00326.
 GN YP00326.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_Taxid=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RA MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moutie S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 DR EMBL; AL14142; CAC89187.1; -
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 104 AA; 12198 MW; 61E90147CEA2CB35 CRC64;

Query Match 2.2%; Score 8; DB 16; Length 104;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 FLDETRNL 309
 DB 44 FLDETRNL 51

RESULT 3

ID 08DI1 PRELIMINARY; PRT; 105 AA.
 AC 08DI1;
 DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN Y0583.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_Taxid=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,

RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Felthous J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM."
 RL J. Bacteriol. 184:4601-4611(2002).
 DR EMBL; AE013660; AAM84171.1; -
 KW Hypothetical protein.
 SO SEQUENCE 105 AA; 12329 MW; 3B7CF288F2515E9C CRC64;

Query Match 2.2%; Score 8; DB 16; Length 105;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 FLDETRNL 309
 DB 45 FLDETRNL 52

RESULT 4

ID 09HS05 PRELIMINARY; PRT; 205 AA.
 AC 09HS05;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE Vng0463c.
 GN VNG0463c.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_Taxid=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sdrooga J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leitner B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Madocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Edhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AE005000; AAC19003.1; -
 KW Complete proteome.
 SO SEQUENCE 205 AA; 22058 MW; C48B843B48AE7562 CRC64;

Query Match 2.2%; Score 8; DB 17; Length 205;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 VHGCGLL 231
 DB 158 VHGCGLL 165

RESULT 5

ID 08XG3 PRELIMINARY; PRT; 299 AA.
 AC 08XG3;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Transcriptional activator of nhA (lysr family) (Transcriptional
 activator protein Nhar).
 GN Nhar OR STM0040 OR STY0048.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=602, 601;

```

RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McCelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogsh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL: AE008695; AAL19004.1; -
DR EMBL: AL627265; CAD01195.1; -
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR000847; HTH_LysR.
DR Pfam: PF00126; HTH_1; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
KW Complete proteome.
SQ SEQUENCE 299 AA; 33998 MW; 49FF81266689D2EE CRC64;

Query Match 2.2%; Score 8; DB 16; Length 299;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 343 GOIKALEE 350
DB 39 GOIKALEE 46

RESULT 6
O81PT9 PRELIMINARY; PRT; 361 AA.
AC O81PT9.
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE CG10512-PC.
GN CG10512.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amenatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.S.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Abmayyan A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

```

```

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster E., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Idegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy C., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wattaman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amenatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreanek D., Farfan D.,
RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Idegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svrtkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnicka F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Flybase;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
EMBL: AE003593; AA012159.1; -
SQ SEQUENCE 361 AA; 37955 MW; 6E92BBE95564F0AE CRC64;

Query Match 2.2%; Score 8; DB 5; Length 361;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 FLDMATT 187
DB 180 FLDMATT 187

```

Db 177 FLDMMAT 184

RESULT 7

097LJ4 PRELIMINARY; PRT; 369 AA.

ID 097LJ4; AC 097LJ4; DT 01-OCT-2001 (TREMUREL. 18, Created) DT 01-OCT-2001 (TREMUREL. 18, Last sequence update) DT 01-DEC-2001 (TREMUREL. 19, Last annotation update) DE Malate dehydrogenase. GN CAC0566. OS Clostridium acetobutylicum. OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium. NCBI_TaxID=1488; RN [1] RP SEQUENCE FROM N.A. RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787; RX MEDLINE-21359325; PubMed-11466286; RA Noelling J., Bleton G., Omeichenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I., Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.; RT Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum. RL J. Bacteriol. 183:4823-4838(2001). DR EMBL: AE007571; AAK78545.1; F. InterPro: IPR003767; Idh_2. DR Pfam: PF02615; Idh_2; 1. KW complete proteome. SQ SEQUENCE 369 AA; 40989 MW; 589826988696C58 CRC64;

Query Match 2.2%; Score 8; DB 16; Length 369; Best Local Similarity 100.0%; Pred. No. 16; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 GGLPLUG 235
Db 232 GGLPLUG 239

RESULT 8

062128 PRELIMINARY; PRT; 372 AA.

ID 062128; AC 062128; DT 01-AUG-1998 (TREMUREL. 07, Created) DT 01-NOV-1998 (TREMUREL. 08, Last sequence update) DT 01-MAR-2003 (TREMUREL. 23, Last annotation update) DE F36A2.3 protein. GN F36A2.3. OS Caenorhabditis elegans. OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderae; Caenorhabditis. NCBI_TaxID=6239; RN [1] RP SEQUENCE FROM N.A. RA Lennard N.; RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases. RN [2] RP SEQUENCE FROM N.A. MEDLINE-94150718; PubMed-7906398; RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R., Smalton N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.; RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans. Nature 368:32-38(1994).

RL Nature 368:32-38(1994).

RN [3] RP SEQUENCE FROM N.A. RA Lennard N.; RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases. DR EMBL: Z96047; CAB09417.1; JOINED. DR EMBL: Z81077; CAB09417.1; JOINED. DR EMBL: Z81077; CAB03073.1; JOINED. DR EMBL: Z96047; CAB03073.1; JOINED. DR WormPep: F36A2.3; CE15964. DR InterPro: IPR003767; Idh_2. DR Pfam: PF02615; Idh_2; 1. SQ SEQUENCE 372 AA; 40461 MW; 69297745F266D80A CRC64;

Query Match 2.2%; Score 8; DB 5; Length 372; Best Local Similarity 100.0%; Pred. No. 16; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 GOCFVAID 287
Db 285 GOCFVAID 292

RESULT 9

Q81PU0 PRELIMINARY; PRT; 381 AA.

ID Q81PU0; AC Q81PU0; DT 01-MAR-2003 (TREMUREL. 23, Created) DT 01-MAR-2003 (TREMUREL. 23, Last sequence update) DT 01-MAR-2003 (TREMUREL. 23, Last annotation update) DE CG10512-PB. GN CG10512. OS Drosophila melanogaster (Fruit fly). OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. NCBI_TaxID=7227; RN [1] RP SEQUENCE FROM N.A. RX MEDLINE-20196006; PubMed-10731132; RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazee R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abell J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brostein P., Brotler P., Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Idagaga C., Jaitai M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Mishina N.V., Mobbry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert C., Remington K., Saunders R.D., Scheller F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spter E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Teator C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 [2]
 RN SEQUENCE FROM N.A.
 RP Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Bazon J., An H., Baldwin D., Bazon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Honck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirski R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RP Mirra S., Crosby M.A., Matthews B.B., Bayraktaroglu I., Campbell K.,
 RA Hradecky P., Huang X., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn B., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome."
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RA Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [5]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AE003593; AAN12158.1;
 SQ SEQUENCE 381 AA; 40221 MW; A847E9F3E0400331 CRC64;
 Query Match 2.2%; Score 8; DB 5; Length 381;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 180 FLDMATT 187
 DB 197 FLDMATT 204
 RESULT 10
 O8KJC9 PRELIMINARY; PRT; 416 AA.
 AC O8KJC9;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Probable sarcosine oxidase beta subunit protein.
 GN SOXB1.
 OS *Rhizobium loti* (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCB1_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R7A;
 RX MEDLINE=2199972; PubMed=12003951;
 RA Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J.,
 RA Brown S.D., Elliott R.M., Fleetwood D.J., McCallum N.G., Rossbach U.,

RA Stuart G.S., Weaver J.E., Webby R.J., de Bruijn F.J., Ronson C.W.;
 RT "Comparative loci analysis of the symbiosis island of
 RT *Mesorhizobium loti* strain R7A."
 RL J. Bacteriol. 184:3086-3095(2002).
 DR EMBL; AL672113; CAD31642.1;
 DR InterPro; IPR006278; SOXB.
 DR TIGRfams; TIGR01373; SOXB; 1.
 SQ SEQUENCE 416 AA; 45286 MW; D77F4C4877ACB707 CRC64;
 Query Match 2.2%; Score 8; DB 2; Length 416;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 LAKERGVA 115
 DB 50 LAKERGVA 57
 RESULT 11
 O961U7 PRELIMINARY; PRT; 446 AA.
 AC O961U7;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE GH06154p.
 GN CG10512.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY047551; AAK77283.1;
 DR FlyBase; FBgn0037057; CG10512.
 DR InterPro; IPR003767; ldb_2.
 DR Pfam; PF02615; ldb_2; 1.
 SQ SEQUENCE 446 AA; 47001 MW; 542A09F12BEE2207 CRC64;
 Query Match 2.2%; Score 8; DB 5; Length 446;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 180 FLDMATT 187
 DB 262 FLDMATT 269
 RESULT 12
 Q9VP68 PRELIMINARY; PRT; 446 AA.
 AC Q9VP68;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG10512 protein.
 GN CG10512.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;

RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA April J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshirei A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palenett K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter R., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley R.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03593; AAF51687.1; -;
 DR FLYBase: FBgn0037057; CG10512.
 DR InterPro: IPR003767; Idh_2.
 DR Pfam: PF02615; Idh_2; 1.
 SQ SEQUENCE 446 AA; 47074 MW; B56F1DA06EEF721D CRC64;

Query Match 2.2%; Score 8; DB 5; Length 446;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 180 FLIDMAT 187
 Db 262 FLIDMAT 269

RESULT 13

O8GGM7 PRELIMINARY; PRT; 474 AA.
 AC O8GGM7;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Antidiotic efflux protein.
 GN LMY.
 OS Streptomyces atroolivaceus.
 OC Bacteria: Actinobacteria: Actinobacteridae: Actinomycetales;
 OC Streptomycineae: Streptomycetaceae: Streptomyces.
 OC NCBI_TaxID=66869;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-2236326; PubMed-1244651;
 RA Cheng Y.O., Tang G.L., Shen B.;
 RT "Identification and localization of the Gene Cluster Encoding Biosynthesis of the Antitumor Macrolactam Telamycin in Streptomyces

RT atroolivaceus S-140.";
 RL J. Bacteriol. 184:7013-7024(2002).
 DR EMBL: AF484556; AAN85538.1; -;
 SQ SEQUENCE 474 AA; 48914 MW; 18E7C82B068EBD5 CRC64;

Query Match 2.2%; Score 8; DB 2; Length 474;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 228 GGLPLGCG 235
 Db 74 GGLPLGCG 81

RESULT 14

O8XK17 PRELIMINARY; PRT; 523 AA.
 AC O8XK17;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein CPE1408.
 GN CPE1408
 OS Clostridium perfringens.
 OC Bacteria: Firmicutes: Clostridia: Clostridiales: Clostridiaceae;
 OC Clostridium.
 OC NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-13 / Type A;
 RX PubMed-11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL: AP003190; BAB8114.1; -;
 DR InterPro: IPR001807; Cl-channel_volt.
 DR InterPro: IPR001891; Na/diCO_sympor.
 DR InterPro: IPR006037; TrkAC.
 DR Pfam: PF02080; TrkA-C; 1.
 DR Pfam: PF00654; voltage_CIC; 1.
 DR PRINTS: PR00762; ClCHANNEL.
 DR PRINTS: PR00173; EDTRNSPORT.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 523 AA; 56185 MW; 0496EA3090C6177C CRC64;

Query Match 2.2%; Score 8; DB 16; Length 523;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 LIGAVGN 99
 Db 339 LIGAVGN 346

RESULT 15

O8VX51 PRELIMINARY; PRT; 651 AA.
 AC O8VX51;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 23, Last annotation update)
 DE Putative receptor-like serine-threonine protein kinase.
 GN PRK-3.
 OS Solanum tuberosum (Potato).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
 OC Asteridae: lamiales; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-cv. Bintje;

RA Montesano M., Kolv V., Maee A., Palva T.;
RT "Novel receptor-like kinases induced by Erwinia carotovora and short
RT oligogalacturonides in potato";
RL Mol. Plant Pathol. 2:339-346(2002).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AJ306628; CAC84518.1; -.
DR InterPro; IPR002902; DUF26.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF01657; DUF26; 2.
DR Pfam; PF00069; pk_kinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 651 AA; 72122 MW; A10D781BCDA7D31E CRC64;

Query Match 2.2%; Score 8; DB 10; Length 651;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 VELADCRG 201
|||||||
Db 85 VELADCRG 92

Search completed: July 24, 2003, 21:27:53
Job time : 107 secs

THIS PAGE BLANK (USPTO)